

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:44:15 ; search time 34.4229 seconds  
(without alignments)  
225.942 Million cell updates/sec

Title: US-09-423-100-1  
Perfect score: 260  
Sequence: 1 MFPTIPLRLFDNMLRAHR.....QEEFEYIPKQYKSYFLQNP 49  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum hit seq length: 3  
Maximum hit seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*  
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	260	100.0	49	AA142855	Human growth hormo
2	260	100.0	92	AA142856	Human growth hormo
3	260	100.0	107	AA142860	hGH-mini-proinsulin
4	260	100.0	134	AA192265	Human anti-angioge
5	260	100.0	140	AA191041	Human growth hormo
6	260	100.0	150	AA142861	Chimeric protein,
7	260	100.0	192	AA190129	Human growth hormo
8	260	100.0	192	AA192264	Human anti-angioge
9	260	100.0	261	AA191299	Human nerve growth

10	260	100.0	262	7	AA161033	Human beta-nerve q
11	260	100.0	262	12	AA11740	Human growth hormo
12	260	100.0	310	11	AA103255	Fusion protein of
13	257	96.8	344	11	AA105313	Segment of B-cell
14	256	98.5	204	23	AA177327	Human growth hormo
15	255	98.1	138	9	AA111226	Sequence of protei
16	255	98.1	179	23	AA147922	Human GH-V seq 10
17	255	98.1	191	7	AA160116	Sequence of human
18	255	98.1	191	18	AA102010	Protein sequence o
19	255	98.1	191	19	AA171289	Human growth hormo
20	255	98.1	191	20	AA175809	Primary amino acid
21	255	98.1	191	20	AA164196	Natural human 22KD
22	255	98.1	191	20	AA104197	Mutant human 22KD
23	255	98.1	191	21	AA178425	Human growth hormo
24	255	98.1	191	22	AA117485	Human growth hormo
25	255	98.1	191	22	AA117486	Human growth hormo
26	255	98.1	191	23	AA144855	Human growth hormo
27	255	98.1	191	23	AA144856	Human growth hormo
28	255	98.1	191	23	AA144857	Human growth hormo
29	255	98.1	191	23	AA144858	Human growth hormo
30	255	98.1	191	23	AA144859	Human growth hormo
31	255	98.1	191	23	AA144860	Human growth hormo
32	255	98.1	191	23	AA144861	Human growth hormo
33	255	98.1	191	23	AA144862	Human growth hormo
34	255	98.1	191	23	AA144863	Human growth hormo
35	255	98.1	191	23	AA144864	Human growth hormo
36	255	98.1	191	23	AA144865	Human growth hormo
37	255	98.1	191	23	AA144866	Human growth hormo
38	255	98.1	191	23	AA144867	Human growth hormo
39	255	98.1	191	23	AA144868	Human growth hormo
40	255	98.1	191	23	AA144869	Human growth hormo
41	255	98.1	191	23	AA144870	Human growth hormo
42	255	98.1	191	23	AA144871	Human growth hormo
43	255	98.1	191	23	AA144872	Human growth hormo
44	255	98.1	191	23	AA144873	Human growth hormo
45	255	98.1	191	23	AA144874	Human growth hormo

ALIGNMENTS

RESULT 1  
ID AAY42855 standard; protein: 49 AA.  
AC AAY42855;  
DT 19-JAN-2000 (first entry)  
DE Human growth hormone (hGH) N-terminal fragment B1.  
XX Growth hormone; chaperone; intramolecular; insulin; precursor;  
XX folding; conformation; chimeric protein; cleavable; recombinant;  
XX production; yield.  
XX  
XX Homo sapiens.  
XX  
XX WO9950302-A1.  
XX  
XX 07-OCT-1999.  
XX  
XX 31-MAR-1998; 98WO-CM00052.  
XX  
XX 31-MAR-1998; 98WO-CM00052.  
XX (TONG-) TONGHUA GANTECH BIOTECHNOLOGY LTD.  
XX Gan Z;  
XX WPI: 1999-610839/52.  
XX New chimeric proteins containing human growth hormone fragment, used  
PT particularly for the production of human insulin.

XX PS Claim 4; Page 28; 46pp; English.

XX CC This sequence represents an N-terminal fragment of human growth

XX CC hormone (hGH) which is a component of a chimeric protein,

XX CC hGH-mini-proinsulin (AA42860). The hGH portion of the chimeric protein

XX CC acts as an intramolecular chaperone (IMC) for the insulin precursor,

XX CC enabling it to fold correctly. A cleavable peptide linker with a

XX CC C-terminal Arg residue (AA42857) enables the hGH portion of the

XX CC chimeric protein to be removed after folding has taken place. Production

XX CC of recombinant human insulin via an hGH-proinsulin chimeric protein can

XX CC provide human insulin with correctly linked cysteine bridges with

XX CC fewer necessary procedural steps, and hence resulting in a higher yield

XX CC of human insulin. The IMC sequences not only protect insulin sequences

XX CC from intracellular degradation by a microorganism host, but also promote

XX CC the folding of the fused insulin precursor, facilitate the solubility of

XX CC the fusion protein and decrease the intermolecular interactions among

XX CC the fusion proteins, thus allowing folding of the fused insulin precursor

XX CC at commercially useful high concentrations. The procedural steps of

XX CC cyanogen bromide cleavage, oxidative sulphydrololysis and related

XX CC purification steps can thus be eliminated, along with the use of high

XX CC concentrations of mercaptan or the use of hydrophobic absorbent resins.

XX SQ Sequence 49 AA:

Query Match 100.0%; Score 260; DB 20; Length 49;

Best Local Similarity 100.0%; Pred. No. 2; 3e-25;

Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MEPTPLSLFNFAMLAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNP 49

DB 1 MEPTPLSLFNFAMLAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNP 49

## RESULT 2

AA42856

ID AA42856 standard; protein; 92 AA.

XX AC AA42856;

XX DT 19-JAN-2000 (first entry)

XX DE Human growth hormone (hGH) N-terminal fragment #2.

XX KW Growth hormone; chaperone; intramolecular; insulin; precursor;

XX KW folding; conformation; chimeric protein; cleavable; recombinant;

XX KW production; yield.

XX OS Homo sapiens.

XX PN WO9950302-A1.

XX PD 07-OCT-1999.

XX PF 31-MAR-1998; 98WO-CN00052.

XX PR 31-MAR-1998; 98WO-CN00052.

XX PA (TONG) TONGHUA GANTECH BIOTECHNOLOGY LTD.

XX PI Gan Z;

XX WPI: 1999-610839/52.

XX PT New chimeric proteins containing human growth hormone fragment, used

XX PT particularly for the production of human insulin.

XX PS Claim 5; Page 28; 46pp; English.

XX CC This sequence represents an N-terminal fragment of human growth hormone

XX CC (hGH) which is a component of a chimeric protein (AA42860) which also

XX CC contains a human insulin precursor (AA42859). The hGH portion of the

XX CC chimeric protein acts as an intramolecular chaperone (IMC) for the

CC insulin precursor, enabling it to fold correctly. A cleavable peptide

CC linker with a C-terminal Arg residue (AA42857) enables the hGH portion

CC of the chimeric protein to be removed after folding has taken place.

CC Production of recombinant human insulin via an hGH-proinsulin chimeric

CC protein can provide human insulin with correctly linked cysteine bridges

CC with fewer necessary procedural steps, and hence resulting in a higher

CC yield of human insulin. The IMC sequences not only protect insulin

CC sequences from intracellular degradation by a microorganism host, but

CC also promote the folding of the fused insulin precursor, facilitate the

CC solubility of the fusion protein and decrease the intermolecular

CC interactions among the fusion proteins, thus allowing folding of the

CC fused insulin precursor at commercially useful high concentrations. The

CC procedural steps of cyanogen bromide cleavage, oxidative sulphydrololysis

CC and related purification steps can thus be eliminated, along with the

CC use of high concentrations of mercaptan or the use of hydrophobic

XX absorbent resins.

SQ Sequence 92 AA:

Query Match 100.0%; Score 260; DB 20; Length 92;

Best Local Similarity 100.0%; Pred. No. 4; 6e-25;

Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MEPTPLSLFNFAMLAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNP 49

DB 1 MEPTPLSLFNFAMLAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNP 49

## RESULT 3

AA42860

ID AA42860 standard; protein; 107 AA.

XX AC AA42860;

XX DT 19-JAN-2000 (first entry)

XX DE hGH-mini-proinsulin chimeric protein.

XX KW Insulin; precursor; growth hormone; chaperone; intramolecular;

XX KW folding; conformation; chimeric protein; cleavable; recombinant;

XX KW production; yield.

XX OS Synthetic.

XX OS Homo sapiens.

XX PN WO9950302-A1.

XX PD 07-OCT-1999.

XX PF 31-MAR-1998; 98WO-CN00052.

XX PR 31-MAR-1998; 98WO-CN00052.

XX PA (TONG) TONGHUA GANTECH BIOTECHNOLOGY LTD.

XX PI Gan Z;

XX WPI: 1999-610839/52.

XX PT New chimeric proteins containing human growth hormone fragment, used

XX PT particularly for the production of human insulin.

XX PS Claim 13; Page 30; 46pp; English.

XX CC This sequence represents a chimeric protein, hGH-mini-proinsulin.

XX CC This chimeric protein contains an N-terminal fragment of human growth

XX CC hormone (hGH) of the sequence given in AA42855, a cleavable peptide

XX CC linker (AA42857), and a human insulin precursor comprising insulin

XX CC A and B chains (AA42859). The hGH portion of the chimeric protein acts

XX CC as an intramolecular chaperone (IMC) for the insulin precursor,

XX CC enabling it to fold correctly. The cleavable peptide linker has a

XX CC C-terminal Arg residue which enables the hGH portion of the

XX CC chimeric protein to be removed after folding has taken place. Production

of recombinant human insulin via an hGH-proinsulin chimeric protein can provide human insulin with correctly linked cysteine bridges with fewer necessary procedural steps, and hence resulting in a higher yield of human insulin. The IMC sequences not only protect insulin sequences from intracellular degradation by a microorganism host, but also promote the folding of the fused insulin precursor, facilitate the solubility of the fusion protein and decrease the intermolecular interactions among the fusion proteins, thus allowing folding of the fused insulin precursor at commercially useful high concentrations. The procedural steps of cyanoogen bromide cleavage, oxidative sulphydrololysis and related purification steps can thus be eliminated, along with the use of high concentrations of mercaptan or the use of hydrophobic absorbent resins.

Sequence 107 AA:  
Query Match 100.0%; Score 260; DB 20; Length 107;  
Best Local Similarity 100.0%; Pred. No. 5.4e-25;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MFPTIPLSRLFNAMLRHRLHJLAFTYQFFEEAYIFKPKYSEFQNP 49  
DB 1 MFPTIPLSRLFNAMLRHRLHJLAFTYQFFEEAYIFKPKYSEFQNP 49

# RESULT 4

AAW92265 standard; Protein: 134 AA.  
AAW92265;  
08-JUN-1999 (first entry)  
Human anti-angiogenic peptide 16K hGH Met-Iprol33.  
Human: anti-angiogenic; prolactin; placental lactogen; hPL; angiogenesis; growth hormone; hGH; hGH-V; capillary endothelial cell proliferation; placental vasculatisation; pregnancy; treatment; angiogenic disease; tumour; inhibitor; malignant; angiofibroma; arteriovenous malformation; arthritis; atherosclerotic plaques; corneal graft neovascularisation; wound healing; proliferative retinopathy; macular degeneration; trachoma; granulation; glaucoma; ocular; uveitis; culture; Osler-Weber syndrome; psoriasis; fibroplasia; scleroderma; Kaposi's sarcoma; vascular adhesion; ulcer; leukaemia; reproductive disorder; contraceptive agent; gene therapy; pre-eclampsia; intrauterine growth retardation; placental dysfunction.  
Homo: sapiens.  
W00451423 A1.  
19 NOV 1998.  
12 MAY 1998; 96WO-US09691.  
13-MAY-1997; 97US-0046394.  
(REPT) UNIV CALIFORNIA.  
Marshall JA, Struman I, Taylor K, Weiner Rf;  
WPI: 1999-045192/04.  
N-PSOR: AAK31707.  
New anti-angiogenic peptides comprise N-terminal fragments of human placental lactogen, human growth hormone, growth hormone variant or human prolactin  
Claim 4: Page 49-50: 87pp: English.  
This invention describes novel human anti-angiogenic peptides derived from 10 to 150 consecutive amino acids selected from the N-terminal end of human placental lactogen (hPL), human growth hormone (hGH), growth hormone variant (hGH-V), or human prolactin. Such peptides (i) inhibit

capillary endothelial cell proliferation and organisation (ii) inhibit angiogenesis in chick chorioallantoic membrane and (iii) binds to at least one specific receptor which does not bind an intact full length hGH, hPL, prolactin or hGH-V. The invention also describes a method for diagnosing a probable abnormality of placental vascularisation during pregnancy. The peptides can be used for treating an angiogenic disease in a subject, for inhibiting tumour formation or growth in a patient, or for modulating vascularisation of a patient's placenta. In particular, the peptides can be used for preventing or treating e.g. malignant tumours, angiofibroma, arteriovenous malformation, arthritic such as rheumatoid arthritis, atherosclerotic plaques, corneal graft neovascularisation, delayed wound healing, proliferative retinopathy such as diabetic retinopathy, macular degeneration, granulations such as those occurring in haemophiliac joints, inappropriate vascularisation in wound healing such as hypertrophic scars or keloid scars, neovascular glaucoma, ocular tumour, uveitis, non-union fractures, Osler-Weber syndrome, psoriasis, pyogenic glaucoma, retrolental fibroplasia, scleroderma, solid tumours, Kaposi's sarcoma, trachoma, vascular adhesions, chronic varicose ulcers, leukaemia, and reproductive disorders such as follicular and luteal cysts and chorioepithelioma. They can also be used as contraceptive agents. DNA encoding the peptides can be used in gene therapy. The measurement of abnormal levels of N-terminal fragments of hGH, hGH-V, prolactin or hPL can be used in assays for impairment of vascular development associated with pre-eclampsia, intrauterine growth retardation, and placental dysfunction.

Sequence 134 AA:

Query Match 100.0%; Score 260; DB 20; Length 134;  
Best Local Similarity 100.0%; Pred. No. 7e-25;  
Matches 49; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 MFPTIPLSRLFNAMLRHRLHJLAFTYQFFEEAYIFKPKYSEFQNP 49  
DB 1 MFPTIPLSRLFNAMLRHRLHJLAFTYQFFEEAYIFKPKYSEFQNP 49

RESULT 5  
AAP91041  
ID AAP91041 standard; protein: 140 AA.  
AAP91041;  
AC AAP91041;  
XX 14-DEC-1989 (first entry).  
XX Human growth hormone segment.  
XX Human growth hormone; fusion protein; thrombin;  
XX geriatric dementia; nervous disorders; human nerve factor.  
XX Homo sapiens (human).  
XX EP329175-A.  
XX 23-AUG-1989.  
XX 17-FEB-1989; 89EP-0102795.  
XX 19-FEB-1988; 88JP-0035042.  
XX (TOYU) TOSOH CORP.  
XX Ohtsuka F;  
XX WPI: 1989-243092/14.  
XX New human nerve growth factor gene encoding fusion protein  
XX - having cleavage site for thrombin, useful for treating geriatric  
XX dementia, etc.  
XX Disclosure: page 21: 38pp: English.  
XX Human growth hormone segment, used at the N terminal of a fusion

CC protein, which contains a thrombin recognition site, and human beta nerve  
 CC growth factor (beta-NGF) at the C-terminal. Beta-NGF can be used to  
 CC control geriatric dementia and other nervous disorders, and can be  
 CC released from the fusion protein by incubation with thrombin (see  
 CC AAN90577-8, AAP91034, AAP91299).

XX Sequence 140 AA;

Query Match 100.0%; Score 260; DB 10; Length 140;  
 Best Local Similarity 100.0%; Pred. No. 7.9e-25;  
 Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEPTPLSLRFDNMLRAHRLHQLAFDTYQFEFEAYIPKEOKYSFLONP 49  
 |||||  
 Db 1 MEPTPLSLRFDNMLRAHRLHQLAFDTYQFEFEAYIPKEOKYSFLONP 49

# RESULT 6

AAV4286;  
 ID AAV42861 standard; protein: 150 AA.

XX  
 AC AAV42861;

DT 19 JAN-2000 (first entry)

XX Chimeric protein, SEQ ID 7.

DE Insulin; precursor: growth hormone; chap- o; intramolecular;  
 XX folding; conformation; chimeric protein; avable; recombinant;  
 KW production; yield.

XX Synthetic.  
 OS Homo sapiens.

XX WO9950302-A1.

XX 07-JUN-1999.

XX 31 MAR-1998; 98WO-CN00052.

XX 31 MAR-1998; 98WO-CN00052.

XX (TUNGHUA GANTECH BIOTECHNOLOGY LTD.

XX Gan Z;

XX WPI: 1999.610839/52.

XX New chimeric proteins containing human growth hormone fragment, used  
 PT particularly for the production of human insulin.

XX Claim 14; Page 30-31; 46pp; English.

XX This sequence represents a chimeric protein, which contains an  
 CC N-terminal fragment of human growth hormone (hGH) of the sequence given  
 CC in AAY42856, a cleavable peptide linker (AAY42857), and a human insulin  
 CC precursor comprising insulin A and B chains (AAY42859). The hGH portion  
 CC of the chimeric protein acts as an intramolecular chaperone (IMC) for  
 CC the insulin precursor, enabling it to fold correctly. The cleavable  
 CC peptide linker has a C-terminal Arg residue which enables the hGH portion  
 CC of the chimeric protein to be removed after folding has taken place.  
 CC Production of recombinant human insulin via an hGH-proinsulin chimeric  
 CC protein can provide human insulin with correctly linked cysteine bridges  
 CC with fewer necessary procedural steps, and hence resulting in a higher  
 CC yield of human insulin. The IMC sequences not only protect insulin  
 CC sequences from intracellular degradation by a microorganism host, but  
 CC also promote the folding of the fused insulin precursor, facilitate the  
 CC solubility of the fusion protein and decrease the intermolecular  
 CC interactions among the fusion proteins, thus allowing folding of the  
 CC fused insulin precursor at commercially useful high concentrations. The  
 CC procedural steps of cyanogen bromide cleavage, oxidative sulphytolysis  
 CC and related purification steps can thus be eliminated, along with the use  
 CC of high concentrations of mercaptan or the use of hydrophobic absorbent

CC resins.

XX Sequence 150 AA;

Query Match 100.0%; Score 260; DB 20; Length 150;  
 Best Local Similarity 100.0%; Pred. No. 7.9e-25;  
 Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEPTPLSLRFDNMLRAHRLHQLAFDTYQFEFEAYIPKEOKYSFLONP 49  
 |||||  
 Db 1 MEPTPLSLRFDNMLRAHRLHQLAFDTYQFEFEAYIPKEOKYSFLONP 49

# RESULT 7

AAP90129  
 ID AAP90129 standard; protein: 192 AA.

XX AAP90129;

XX 25 MAR-2004 (updated)

DT 06-FEB-1996 (revised)

DT 01-NOV-1989 (first entry)

XX Human growth hormone.

XX Human growth hormone; fusion protein; recombinant  
 KW vector.

XX Homo sapiens (Human).

XX JP01144981-A.

XX 07-JUN 1989.

XX 02-JUL-1987; 87JP-0304937.

XX 02-DEC-1987; 87JP-0304937.

XX (WAKT) WAKUNAGA SEIVAKU KK.

XX WPI: 1989-209284/29.

XX N-PSDB: AAN90269.

XX Recombinant vector contg. fusion protein consisting of human  
 PT growth hormone or deriv. ligated to foreign protein, for stability  
 PT and high yield.

XX Disclosure: Fig 1: 15pp; Japanese.

XX The invention consists of a vector contg. a fusion protein which is  
 CC formed by ligating, downstream of a promoter, hGH or a deriv. (pref.  
 CC formed by substn. of Met-14 with leu) and a foreign protein.  
 CC Stability of the vector in the host is greatly increased so the  
 CC protein yield is higher.  
 CC (Updated on 25-MAR-2003 to correct PA field.)

XX Sequence 192 AA;

Query Match 100.0%; Score 260; DB 10; Length 192;  
 Best Local Similarity 100.0%; Pred. No. 1e-24;  
 Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEPTPLSLRFDNMLRAHRLHQLAFDTYQFEFEAYIPKEOKYSFLONP 49  
 |||||  
 Db 1 MEPTPLSLRFDNMLRAHRLHQLAFDTYQFEFEAYIPKEOKYSFLONP 49

# RESULT 8

AAW92264  
 ID AAW92264 standard; protein: 192 AA.

XX AAW92264;

AC AAW92264;



XX Human beta-nerve growth factor gene product.  
 DE Beta-NGF; E.coli; ds.  
 XX Homo sapiens.  
 OS  
 XX Key Location/Qualifiers  
 FH Protein 145..262  
 FT  
 XX JF01205485 A.  
 PN  
 XX 11-SEP-1986.  
 PD  
 XX 09-MAR-1985; 85JP-0045773.  
 PF  
 XX 09-MAR-1985; 85JP-0045773.  
 PR  
 XX (OTSU/) OTSUKA E.  
 PA  
 XX WPI: 1986-281646/43.  
 DR  
 XX Gene segment of human nerve growth factor, used in production of  
 PT NGF producing recombinant Escherichia strain.  
 PS  
 XX Claim 32; Page 482; 71pp; Japanese.  
 CC The protein is a direct translation of the upstream tryptophan  
 CC promoter operator lacking its attenuation sequence and human  
 CC beta NGF sequence. The product may be efficiently expressed from a  
 CC transformed E.coli expression system.  
 CC See also AAN60816-7.  
 XX  
 XX Sequence 262 AA;

Query Match 100.0%; Score 260; DB 7; Length 262;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-24;  
 Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MEPTIPLSRLFONAMLRHRLHOLAFDTYOFFEEAYIPKEQKYSFLONP 49  
 DB 1 MEPTIPLSRLFONAMLRHRLHOLAFDTYOFFEEAYIPKEQKYSFLONP 49

RESULT 11  
 AAR1174:  
 ID AAR11740 standard; Protein; 262 AA.  
 XX  
 AC AAR11740;  
 XX  
 XX 25-MAR-2003 (updated)  
 DT 25-JUN-1991 (first entry)  
 XX

DE Human growth hormone/human nerve growth factor beta fusion protein.  
 KW hGH; hNGF; nervous system diseases; dementia.  
 XX Homo sapiens.  
 OS  
 XX JF01067598 A.  
 PN  
 XX 22-MAR-1991.  
 PD  
 XX 07-AUG-1989; 89JP-0202835.  
 PF  
 XX 07-AUG-1989; 89JP-0202835.  
 PR  
 XX (T-XT) TOSOH CORP.  
 PA  
 XX WPI: 1991-128768/18.  
 DR N-PSDB; AAR11578.  
 XX  
 XX Purification of human neuron growth factor beta subunit-contg. protein.

PT by contacting with gel having cation exchange gp. in presence of  
 PT urea  
 XX  
 XX Disclosure; fig 1; 7pp; Japanese.  
 PS  
 XX A recombinant human nerve growth factor beta subunit-contg. protein  
 CC can be produced as this fusion protein. It is purified by contacting  
 CC a gel having a cation exchange gp. with the fusion protein, in the  
 CC presence of urea. The purified protein is useful in a medication  
 CC for treating disorders of the nervous system, eg dementia.  
 CC (Updated on 25-MAR-2003 to correct PF field.)  
 XX  
 XX Sequence 262 AA;

Query Match 100.0%; Score 260; DB 12; Length 262;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-24;  
 Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MEPTIPLSRLFONAMLRHRLHOLAFDTYOFFEEAYIPKEQKYSFLONP 49  
 DB 1 MEPTIPLSRLFONAMLRHRLHOLAFDTYOFFEEAYIPKEQKYSFLONP 49

RESULT 12  
 AAR03255:  
 ID AAR03255 standard; Protein; 310 AA.  
 XX  
 AC AAR03255;  
 XX

DT 19-JUL-1990 (first entry)  
 XX  
 XX Fusion protein of B-cell stimulatory factor: 2 and B-cell  
 DE differentiation factor.  
 DE  
 XX B-cell stimulatory factor: 2; interleukin 6; B cell differentiation;  
 KW interleukin-5; fusion protein.  
 XX Homo sapiens.  
 OS

XX JP02013375-A.  
 PN  
 XX 17-JAN-1990.  
 PD  
 XX 01-JUL-1988; 88JP-0162556.  
 PF  
 XX 01-JUL-1988; 88JP-0162556.  
 PR  
 XX (TOYO) TOSOH CORP.  
 PA  
 XX WPI: 1990-062207/04.  
 DR N-PSDB; AAC02028.  
 XX

PT Prepn. of human B-cell differentiation factor - from specified DNA  
 PT sequence segment, by recombinant DNA technique, gives protein of  
 PT specified amino acid sequence.  
 XX  
 XX Claim 31; Page 9; 17pp; Japanese.  
 PS  
 XX The protein is produced by fusing DNA encoding Bb (IL-) with DNA  
 CC encoding BSP-2 (IL-5) and ligating the product into an expression vector  
 CC See also AAR05311 and AAR05313.  
 XX  
 XX Sequence 310 AA;

Query Match 100.0%; Score 260; DB 11; Length 310;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-24;  
 Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MEPTIPLSRLFONAMLRHRLHOLAFDTYOFFEEAYIPKEQKYSFLONP 49  
 DB 1 MEPTIPLSRLFONAMLRHRLHOLAFDTYOFFEEAYIPKEQKYSFLONP 49

```

PI Yoo JG, Song YH;
XX
XX WPI: 2002-185396/24.
XX N-PSDB: AHL55999.
XX
XX Recombinant human growth hormone having collagenase recognition region
XX
XX Disclosure: Fig 3; 8pp; Korean.
XX
XX The invention relates to recombinant human growth hormone having a
XX collagenase recognition region.
XX
XX Sequence 204 AA:
SQ
    Query Match          98.5%; Score 256; DB 23; Length 204;
    Best Local Similarity 98.0%; Pred. No. 3.5e-24;
    Matches 48; Conservative 1; Mismatches 0; Indels 0; Gaps
    QY 1 MEPTPLSLSCFENAMRAHRLHQLAFDTYQFFEEAYIPKEOKYSFLQNP 49
    QH GIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
    QH IVVFPLSLSLCENAMRAHRLHQLAFDTYQFFEEAYIPKEOKYSFLQNP 61
    QH
RESULT 15
AAP81226
ID AAP81226 standard: [protein; 138 AA.
AC
AC AAP81226;
XX
XX 25-MAR-2003 (updated)
D:
D: 20-NOV-1990 (first entry)
XX
XX Sequence of protein with somatomedin-like activity.
XX
XX Growth hormone.
XX
XX Synthetic.
OS
OS JP63167798-A.
XX
XX 11-JUL-1988.
PD
XX
XX 29-DEC-1986; 86JP-0310177.
PF
XX
XX 29-DEC-1986; 86JP-0310177.
XX
XX (TOYJ ) TOYO SUGA MFG CO LTD.
PA
XX
XX WPI: 1988 232632/33.
XX N-PSDB: AAP81605.
XX
XX Polypeptide with somatomedin-like activity -
XX by culturing bacterium transformed by plasmid contg. gene
XX segment with specified DNA sequence
XX
XX Claim 2(1): Page 609; 9pp; Japanese.
XX
XX The polypeptide (AAP81226) with somatomedin-like activity and the DNA
XX (AAP81605) encoding it are claimed. A Met residue: gp. may be added to
XX the N-terminal. The polypeptide acts on the bone structure of mammals,
XX including humans, to promote bone growth. The polypeptide has high
XX production rate and is easily extracted from bacterial culture medium
XX and refined for use as a bone growth accelerator.
XX (updated on 25-MAR-2003 to correct PA field.)
XX
XX Sequence 138 AA:
SQ
    Query Match          98.1%; Score 255; DB 9; Length 138;
    Best Local Similarity 100.0%; Pred. No. 3e-24;
    Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps
    QY 2 FPTPLSLFDNAMRAHRLHQLAFDTYQFFEEAYIPKEOKYSFLQNP 49
    QH

```

DD 1 FPTIPLSLFDNMLRAHRLHQLAFDTVOFFEEAYIPKEQKYSFLQNP 48

Search completed: September 15, 2003, 12:00:56  
Job time : 35.4229 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:57:40 : Search time 11.2401 Seconds  
(without alignments)  
184.449 Million cell updates/sec

Title: US-09-423-100-1

Perfect score: 260

Sequence: 1 MEPTPLSRLEFONAMLRARR.....OFFEEAYIPKEQKYSFLQNP 49

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 428717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 428717

Minimum db seq length: 0

Maximum db seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2.6/plodata/1/1aa/5A MB.pep:\*  
2: /cgn2.6/plodata/1/1aa/5B MB.pep:\*  
3: /cgn2.6/plodata/1/1aa/6A MB.pep:\*  
4: /cgn2.6/plodata/1/1aa/6B MB.pep:\*  
5: /cgn2.6/plodata/1/1aa/PCMB.pep:\*  
6: /cgn2.6/plodata/1/1aa/backtillesl.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	260	100.0	192	1	US-08-093-383-1
2	255	98.1	191	4	US-09-284-878-5
3	255	98.1	194	2	US-08-383-621-4
4	255	98.1	194	3	US-08-459-906-4
5	255	98.1	217	3	US-08-589-028-10
6	255	98.1	217	3	US-08-784-582-10
7	255	98.1	217	3	US-08-785-271-10
8	255	98.1	217	3	US-08-759-628-11
9	255	98.1	217	4	US-09-284-878-1
10	255	98.1	241	4	US-09-424-620B-25
11	255	98.1	245	4	US-09-280-030-66
12	255	98.1	274	3	US-08-784-582-71
13	255	98.1	360	3	US-08-784-582-73
14	249	95.8	191	4	US-09-465-461-1
15	249	95.8	217	1	US-08-187-756C-4
16	249	95.8	217	1	US-08-469-486-51
17	249	95.8	217	2	US-08-469-651-51
18	249	95.8	217	2	US-08-710-317-4
19	249	95.8	217	4	US-09-411-657-4
20	248	95.4	191	3	US-08-800-215C-16
21	248	95.4	191	3	US-08-800-215C-18
22	248	95.4	191	3	US-08-800-215C-20
23	248	95.4	400	4	US-09-420-819-37
24	248	95.4	401	4	US-09-420-819-36
25	237	91.2	71	1	US-08-314-586-24
26	233	89.6	70	1	US-07-920-519-24
27	233	89.6	70	3	US-08-115-753-26

28	164.5	63.3	191	1	US-08-458-824-8	Sequence 8, Appli
29	164	63.1	176	3	US-08-791-728-1	Sequence 1, Appli
30	164	63.1	176	4	US-08-990-774-1	Sequence 1, Appli
31	161.5	62.1	191	1	US-07-963-310D-4	Sequence 4, Appli
32	159.5	61.3	190	1	US-08-388-267C-2	Sequence 2, Appli
33	159.5	61.3	190	4	US-09-277-720-2	Sequence 2, Appli
34	159.5	61.3	191	6	5210180-1	Patent No. 5210180
35	159.5	61.3	193	1	US-07-621-197C-2	Sequence 2, Appli
36	159.5	61.3	193	1	US-08-363-982-2	Sequence 1, Appli
37	159.5	61.3	193	2	US-08-381-621-1	Sequence 1, Appli
38	159.5	61.3	193	3	US-08-459-906-1	Sequence 1, Appli
39	159.5	61.3	216	2	US-09-105-651-3	Sequence 1, Appli
40	159.5	61.3	216	2	US-09-105-651-3	Sequence 2, Appli
41	158	60.8	176	3	US-08-791-728-2	Sequence 2, Appli
42	158	60.8	176	4	US-08-990-774-2	Sequence 2, Appli
43	158	60.8	177	1	US-08-187-756C-6	Sequence 6, Appli
44	158	60.8	177	2	US-08-710-324A-6	Sequence 6, Appli
45	158	60.8	177	4	US-09-411-657-5	Sequence 6, Appli

#### ALIGNMENTS

RESULT 1  
US-08-093-383-1  
: Sequence 1, Application US/08093383  
: Patent No. 5489529  
: GENERAL INFORMATION:  
: APPLICANT: DeBoer, Herman A.  
: APPLICANT: Heyncker, Herbert L.  
: TITLE OF INVENTION: DNA for Expression of Bovine Growth Hormone  
: NUMBER OF SEQUENCES: 30  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Genentech, Inc.  
: STREET: 460 Point San Bruno Blvd  
: CITY: South San Francisco  
: STATE: California  
: COUNTRY: USA  
: ZIP: 94080  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: 5.25 inch, 360 kb floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: patin (Genentech)  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/093.383  
: FILING DATE: 14-JUL-1993  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: 07/619827  
: FILING DATE: 28-NOV-1990  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: 07/198824  
: FILING DATE: 05-APR 1988  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: 06/632361  
: FILING DATE: 19-JUL-1984  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: 06/401687  
: FILING DATE: 18-SEP-1981  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Johnston, Sean A.  
: REGISTRATION NUMBER: P35,910  
: REFERENCE/DOCKET NUMBER: 46C4  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 415/225-3562  
: TELEFAX: 415/952-9881  
: TELEX: 910/371-7168  
: INFORMATION FOR SEQ ID NO: 1:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 192 amino acids  
: TYPE: amino acid

```

; TOPOLOGY: Linear
; US-08-054-383-1

Query Match      100.0%; Score 260; ID 1; Length 192;
Best Local Similarity 100.0%; Pred. No. 4,6e-30;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFPTPLSRFLDNAMLAHRLHQLAFDTYQFFEEAYIPKEOKYSFLQNP 49
|||||
DB 1 MFPTPLSRFLDNAMLAHRLHQLAFDTYQFFEEAYIPKEOKYSFLQNP 49
|||||

RESULT 2
US-09-284-878-5
; Sequence 5, Application US/09284878
; Patent No. 6342175
; GENERAL INFORMATION:
; APPLICANT: Olazaran, Martha Guerrero
; APPLICANT: Saldaña, Hugo Parrota
; APPLICANT: Salvado, Jose Maria Viader
; TITLE OF INVENTION: Genetically Modified Methylotrophic P. pastoris Yeast for the
; TITLE OF INVENTION: Production and Secretion of the Human Growth Hormone
; FILE REFERENCE: 1829-60100-06
; CURRENT APPLICATION NUMBER: US/09/284.878
; CURRENT FILING DATE: 1999-07-21
; PRIOR APPLICATION NUMBER: PCT/MX97/00033
; PRIOR FILING DATE: 1997-10-24
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-284-878-5

Query Match      98.18; Score 255; DB 4; Length 191;
Best Local Similarity 100.0%; Pred. No. 2,4e-29;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 FPTPLSKFLDNAMLAHRLHQLAFDTYQFFEEAYIPKEOKYSFLQNP 49
|||||
DB 1 FPTPLSKFLDNAMLAHRLHQLAFDTYQFFEEAYIPKEOKYSFLQNP 48
|||||

RESULT 3
US-08-459-906-4
; Sequence 4, Application US/08384621
; Patent No. 5951972
; GENERAL INFORMATION:
; APPLICANT: Daley, Michael J.
; APPLICANT: Buckwalter, Brian L.
; APPLICANT: Cady, Susan M.
; APPLICANT: Shieh, Hong-Ming
; APPLICANT: Bohlen, Peter
; APPLICANT: Seddon, Andrew P.
; TITLE OF INVENTION: Stabilization of Somatotropins And other
; TITLE OF INVENTION: Proteins By Modification Of Cysteine Residues
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07470-8426
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,906
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Darryl L.
; REGISTRATION NUMBER: 34,276
; REFERENCE/DOCKET NUMBER: 31,278-03
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-831-3247
; TELEFAX: 201-831-3305
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 194 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-459-906-4

```

```

; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/766,142
; FILING DATE: 25-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsevdos, Estelle J.
; REGISTRATION NUMBER: 31,145
; REFERENCE/DOCKET NUMBER: 31,278-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 203-321-2756
; TELEFAX: 203-321-2971
; TELEX: 203-710-474-4059
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 194 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-383-621-4

Query Match      98.18; Score 255; DB 4; Length 194;
Best Local Similarity 100.0%; Pred. No. 2,5e-29;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 FPTPLSKFLDNAMLAHRLHQLAFDTYQFFEEAYIPKEOKYSFLQNP 49
|||||
DB 4 FPTPLSKFLDNAMLAHRLHQLAFDTYQFFEEAYIPKEOKYSFLQNP 51
|||||

RESULT 4
US-08-459-906-4
; Sequence 4, Application US/08384621
; Patent No. 6010995
; GENERAL INFORMATION:
; APPLICANT: Daley, Michael J.
; APPLICANT: Buckwalter, Brian L.
; APPLICANT: Cady, Susan M.
; APPLICANT: Shieh, Hong-Ming
; APPLICANT: Bohlen, Peter
; APPLICANT: Seddon, Andrew P.
; TITLE OF INVENTION: Stabilization of Somatotropins and Other
; TITLE OF INVENTION: Proteins by Modification of Cysteine Residues
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07470-8426
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,906
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Darryl L.
; REGISTRATION NUMBER: 34,276
; REFERENCE/DOCKET NUMBER: 31,278-03
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-831-3247
; TELEFAX: 201-831-3305
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 194 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-459-906-4

```

Query Match 98.1% Score 255; DB 3; Length 194;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-29;  
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 FPTPLSRFDNMLRAHRLHQLAFDTYQEEFAYIPKEQKYSFLQNP 49  
 |||||  
 DB 4 FPTPLSRFDNMLRAHRLHQLAFDTYQEEFAYIPKEQKYSFLQNP 51

RESULT 5  
 US-08-589-028-10  
 : Sequence 10, Application US/08589028  
 : Patent No. 6087129  
 : GENERAL INFORMATION:  
 : APPLICANT: Newgard, Christopher B.  
 : APPLICANT: Halban, Philippe  
 : APPLICANT: No. 6087129mington, Karl D.  
 : APPLICANT: Clark, Samuel A.  
 : APPLICANT: Thiipen, Anice E.  
 : APPLICANT: Quade, Christian  
 : APPLICANT: Kruse, Fred  
 : TITLE OF INVENTION: Recombinant Expression of Proteins from  
 : NUMBER OF SEQUENCES: 50  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Arnold, White & Durkee  
 : STREET: P. O. Box 4433  
 : CITY: Houston  
 : STATE: TX  
 : COUNTRY: USA  
 : ZIP: 77210-4433  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC Compatible  
 : OPERATING SYSTEM: PC-DOS/MS DOS  
 : SOFTWARE: Patent In Release #1.0, Version #1.30  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/589,028  
 : FILING DATE: Concurrently herewith  
 : CLASSIFICATION: 435  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Highlander, Steven L.  
 : REGISTRATION NUMBER: 47,642  
 : REFERENCE/DOCKET NUMBER: UTSD:426\HYL  
 : TELEPHONE: (512) 418-3000  
 : TELEFAX: (512) 474-7577  
 : INFORMATION FOR SEQ ID NO: 10:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 217 amino acids  
 : TYPE: amino acid  
 : STRANDEDNESS:  
 : TOPOLOGY: linear  
 : US-08-589-028-10

Query Match 98.1% Score 255; DB 3; Length 217;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-29;  
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 FPTPLSRFDNMLRAHRLHQLAFDTYQEEFAYIPKEQKYSFLQNP 49  
 |||||  
 DB 27 FPTPLSRFDNMLRAHRLHQLAFDTYQEEFAYIPKEQKYSFLQNP 74

RESULT 6  
 US-08-744-582-10  
 : Sequence 10, Application US/08744582  
 : Patent No. 6110707  
 : GENERAL INFORMATION:  
 : APPLICANT: Newgard, Christopher B.  
 : APPLICANT: Halban, Philippe A.  
 : APPLICANT: No. 6110707mington, Karl D.

APPLICANT: Clark, Samuel A.  
 APPLICANT: Thiipen, Anice E.  
 APPLICANT: Quade, Christian  
 APPLICANT: Kruse, Fred  
 APPLICANT: McGarry, Dennis  
 TITLE OF INVENTION: RECOMBINANT EXPRESSION OF PROTEINS FROM  
 TITLE OF INVENTION: SECRETORY CELL LINES  
 NUMBER OF SEQUENCES: 79  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Arnold, White & Durkee  
 STREET: P. O. Box 4433  
 CITY: Houston  
 STATE: Texas  
 COUNTRY: USA  
 ZIP: 77210  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC Compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/784,582  
 FILING DATE: Concurrently herewith  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/028,427  
 FILING DATE: 15-OCT-1996  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08,389,028  
 FILING DATE: 19-JAN-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Highlander, Steven L.  
 REGISTRATION NUMBER: 37,642  
 REFERENCE/DOCKET NUMBER: UTSD-514  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 512/418-3000  
 TELEFAX: 512/474-7577  
 INFORMATION FOR SEQ ID NO: 10:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 217 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 US-08-784-582-10

Query Match 98.1% Score 255; DB 3; Length 217;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-29;  
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 FPTPLSRFDNMLRAHRLHQLAFDTYQEEFAYIPKEQKYSFLQNP 49  
 |||||  
 DB 27 FPTPLSRFDNMLRAHRLHQLAFDTYQEEFAYIPKEQKYSFLQNP 74

RESULT 7  
 US-08-785-271-10  
 : Sequence 10, Application US/08785271  
 : Patent No. 6194176  
 : GENERAL INFORMATION:  
 : APPLICANT: Newgard, Christopher B.  
 : APPLICANT: Halban, Philippe A.  
 : APPLICANT: No. 6194176mington, Karl D.  
 : APPLICANT: Clark, Samuel A.  
 : APPLICANT: Thiipen, Anice E.  
 : APPLICANT: Quade, Christian  
 : APPLICANT: Kruse, Fred  
 : TITLE OF INVENTION: RECOMBINANT EXPRESSION OF PROTEINS FROM  
 : TITLE OF INVENTION: SECRETORY CELL LINES  
 : NUMBER OF SEQUENCES: 56  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Arnold, White & Durkee  
 : STREET: P. O. Box 4433  
 : CITY: Houston

```

; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/785,271
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/589,028
; FILING DATE: 19-1AN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: 08/589,028
; TELEPHONE: 512/418-3003
; TELEFAX: 512/434-7577
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 217 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-785,271-10

Query Match 98.1%; Score 255; DB 1; Length 217;
Best Local Similarity 100.0%; Pred. No. 2.9e-29;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 FETIPLSRFDNAMLRAHRLHQAFDTYQFEFFAYIPKEOKYSFLGNP 49
DB 27 FETIPLSRFDNAMLRAHRLHQAFDTYQFEFFAYIPKEOKYSFLGNP 74

RESULT 8
US-08-785,271-10
; Sequence 1; Application US/08/59628
; Patent No. 6225446
; GENERAL INFORMATION:
; APPLICANT: Altman, Scott W.
; APPLICANT: Rock, Fernando L.
; APPLICANT: Bazan, J. Fernando
; APPLICANT: Kastelein, Robert A.
; TITLE OF INVENTION: MUTATIONAL VARIANTS OF MAMMALIAN PROTEINS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/759,628
; FILING DATE: 05-DEC-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/008,574
; FILING DATE: 06-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX05520

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-852-9196
; TELEFAX: 415-496-1200
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 217 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 32..53
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 94..115
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 133..154
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 192..216
; OTHER INFORMATION: /note "the peptides above are
; OTHER INFORMATION: depicted in Figure 1"
; US 08-759-628 11

Query Match 98.1%; Score 255; DB 4; Length 217;
Best Local Similarity 100.0%; Pred. No. 2.9e-29;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 FETIPLSRFDNAMLRAHRLHQAFDTYQFEFFAYIPKEOKYSFLGNP 49
DB 27 FETIPLSRFDNAMLRAHRLHQAFDTYQFEFFAYIPKEOKYSFLGNP 74

RESULT 9
US-09-284-878-1
; Sequence 1; Application US/09284878
; Patent No. 6342375
; GENERAL INFORMATION:
; APPLICANT: Olazaran, Martha Guerrero
; APPLICANT: Saldana, Hugo Barrera
; APPLICANT: Salvado, Jose Maria Viader
; TITLE OF INVENTION: Genetically Modified Methylotrophic P. pastoris Yeast for th
; FILE REFERENCE: 1829.0010000
; CURRENT APPLICATION NUMBER: US/09/284,878
; CURRENT FILING DATE: 1999-07-21
; PRIOR APPLICATION NUMBER: PCT/MX97/00033
; PRIOR FILING DATE: 1997-10-24
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 217
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-284-878-1

Query Match 98.1%; Score 255; DB 4; Length 217;
Best Local Similarity 100.0%; Pred. No. 2.9e-29;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 FETIPLSRFDNAMLRAHRLHQAFDTYQFEFFAYIPKEOKYSFLGNP 49
DB 27 FETIPLSRFDNAMLRAHRLHQAFDTYQFEFFAYIPKEOKYSFLGNP 74

RESULT 10
US 09-424-620H-25
; Sequence 25; Application US/09424620H
; Patent No. 6391585
; GENERAL INFORMATION:
; APPLICANT: HANIL SYNTHETIC FIBER CO., LTD.

```

JANG, Ki-Ryong  
MOON, Jae-Woong  
BAE, Cheon-Soon  
YANG, Doo-Suk  
LEE, Jee-Won  
SFONG, Baik-Lin  
TITLE OF INVENTION: Process for preparing recombinant proteins using highly efficient expression vector from *Saccharomyces cerevisiae*  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BACHMAN & LAPOINTE, P.C.  
STREET: Suite 1201, 900 Chapel Street  
CITY: New Haven  
STATE: Connecticut  
COUNTRY: U.S.A.  
ZIP: 06510-2802  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB storage  
COMPUTER: IBM  
OPERATING SYSTEM: WINDOWS 95/NT  
SOFTWARE: MS WORD  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/99/424,100  
FILING DATE: 24-Nov-1999  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 241 amino acids  
TYPE: amino acid  
MOLECULE TYPE: PROTEIN  
SEQUENCE DESCRIPTION: SEQ ID NO: 25:  
US-09-424-620B-25

Query Match 98.1%; Score 255; DB 4; Length 241;  
Best Local Similarity 100.0%; Pred. No. 3.3e-29;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 FPTPLSRFDNAMLRAHRLHQLAFDTYQEFFEAYIPKEQKYSFLQNP 49  
|||||  
DB 51 FPTPLSRFDNAMLRAHRLHQLAFDTYQEFFEAYIPKEQKYSFLQNP 98

RESULT: 11  
US-09-280-030-66  
Sequence 66, Application US/992800030A  
Patent No. 653595  
GENERAL INFORMATION:  
APPLICANT: Sato, Seiji  
APPLICANT: Higashikuni, Naohiko  
APPLICANT: Kado, Toshiyuki  
APPLICANT: Kondo, Masaaki  
TITLE OF INVENTION: DNAS ENCODING NEW FUSION PROTEINS AND PROCESSES FOR  
TITLE OF INVENTION: PREPARING USEFUL POLYPEPTIDES THROUGH EXPRESSION OF THE  
FILE REFERENCE: 382.1026  
CURRENT APPLICATION NUMBER: US/99/280,030A  
EARLIER FILING DATE: 1999-03-26  
EARLIER APPLICATION NUMBER: JP10-87339/1998  
EARLIER FILING DATE: 1998-03-31  
NUMBER OF SEQ ID NOS: 66  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 66  
LENGTH: 245  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Designated is  
OTHER INFORMATION: an amino acid sequence of MWPsp-MWPmp20-TEV-G-CH  
US-09-280-030-66

Query Match 98.1%; Score 255; DB 4; Length 245;  
Best Local Similarity 100.0%; Pred. No. 3.3e-29;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 FPTPLSRFDNAMLRAHRLHQLAFDTYQEFFEAYIPKEQKYSFLQNP 49  
|||||  
DB 55 FPTPLSRFDNAMLRAHRLHQLAFDTYQEFFEAYIPKEQKYSFLQNP 102

RESULT 12  
US-08-784-582-71  
Sequence 71, Application US/08784582  
Patent No. 6110707  
GENERAL INFORMATION:  
APPLICANT: Newgard, Christopher B.  
APPLICANT: Halban, Philippe A.  
APPLICANT: No. 6110707minqton, Karl D.  
APPLICANT: Clark, Samuel A.  
APPLICANT: Thigpen, Anice E.  
APPLICANT: Oseade, Christian  
APPLICANT: Kruse, Fred  
APPLICANT: McGarity, Dennis  
TITLE OF INVENTION: RECOMBINANT EXPRESSION OF PROTEINS FROM  
TITLE OF INVENTION: SECRETORY CELL LINES  
NUMBER OF SEQUENCES: 79  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/784,582  
FILING DATE: Concurrently Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/028,427  
FILING DATE: 15-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/589,028  
FILING DATE: 19-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Highlander, Steven L.  
REGISTRATION NUMBER: 37,642  
REFERENCE/DOCKET NUMBER: UIDS:514  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 512/418-3000  
TELEFAX: 512/474-7577  
INFORMATION FOR SEQ ID NO: 71:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 274 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-784-582-71

Query Match 98.1%; Score 255; DB 3; Length 274;  
Best Local Similarity 100.0%; Pred. No. 3.8e-29;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 FPTPLSRFDNAMLRAHRLHQLAFDTYQEFFEAYIPKEQKYSFLQNP 49  
|||||  
DB 27 FPTPLSRFDNAMLRAHRLHQLAFDTYQEFFEAYIPKEQKYSFLQNP 74

RESULT 13  
US-08-784-582-73  
Sequence 73, Application US/08784582  
Patent No. 6110707  
GENERAL INFORMATION:

APPLICANT: Newgard, Christopher B.  
APPLICANT: Halban, Philippe A.  
APPLICANT: No. 6110707mington, Karl D.  
APPLICANT: Clark, Samuel A.  
APPLICANT: Thigpen, Anice E.  
APPLICANT: Quade, Christian  
APPLICANT: Kruse, Fred  
APPLICANT: McGarry, Dennis  
TITLE OF INVENTION: RECOMBINANT EXPRESSION OF PROTEINS FROM  
TITLE OF INVENTION: SECRETORY CELL LINES  
NUMBER OF SEQUENCES: 79  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P. O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC DOS/MS DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/784,562  
FILING DATE: Concurrently Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/028,427  
FILING DATE: 15-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/589,028  
FILING DATE: 19-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Highlander, Steven L.  
REGISTRATION NUMBER: 37,642  
REFERENCE/DOCKET NUMBER: UTSD:514  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 512/418-3000  
TELEFAX: 512/474-7577  
INFORMATION FOR SEQ ID NO: 73:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 360 amino acids  
TYPE: amino acid  
STRANDEDNESS: linear  
TOPOLOGY:  
US-08-784-562-73

Query Match 98.1% Score 255; DB 3; Length 360;  
Best Local Similarity 100.0%; Pred. No. 5.4e-24;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Caps 0;

OY 2 FPTIPLSRFDNAMLRAHRLHQLAFDTYQFEFFAYIPKEOKYSFLONP 49  
|||||  
DB 27 FPTIPLSRFDNAMLRAHRLHQLAFDTYQFEFFAYIPKEOKYSFLONP 74

## RESULT 14

US-09-465-461-1  
Sequence 1, Application US/09465461  
Patent No. 6148444  
GENERAL INFORMATION:

APPLICANT: CHAPPEL, Scott  
TITLE OF INVENTION: Human Growth Hormone to stimulate hematopoiesis and immune reconst  
TITLE OF INVENTION: after hematopoietic stem cell transplantation in humans  
FILE REFERENCE: CHAPPEL-6.1  
CURRENT APPLICATION NUMBER: US/09/465,461  
PRIOR FILING DATE: 1999-12-17  
PRIOR FILING DATE: 1998-12-17  
NUMBER OF SEQ ID NOS: 1  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1

LENGTH: 191  
TYPE: PRI  
ORGANISM: homo sapiens  
US-09-465-461-1

Query Match 95.8% Score 249; DB 4; Length 191;  
Best Local Similarity 97.9%; Pred. No. 1.8e-28;  
Matches 47; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 2 FPTIPLSRFDNAMLRAHRLHQLAFDTYQFEFFAYIPKEOKYSFLONP 49  
|||||  
DB 1 FPTIPLSRFDNAMLRAHRLHQLAFDTYQFEFFAYIPKEOKYSFLONP 48

## RESULT 15

US-08-187-756C-4  
Sequence 4, Application US/08187756C  
Patent No. 5597709

GENERAL INFORMATION:  
APPLICANT: POSEN, Eli A.  
TITLE OF INVENTION: Human Growth Hormone  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, HAIN, GILFILLAN,  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07068

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/187,756C  
FILING DATE: January 27, 1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: FERRARO, GREGORY D.  
REGISTRATION NUMBER: 36,134  
REFERENCE/DOCKET NUMBER: 325800-55  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-954-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 217 AMINO ACIDS  
TYPE: AMINO ACID  
STRANDEDNESS:  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
US-08-187-756C-4

Query Match 95.8% Score 249; DB 1; Length 217;  
Best Local Similarity 97.9%; Pred. No. 2.1e-24;  
Matches 47; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 2 FPTIPLSRFDNAMLRAHRLHQLAFDTYQFEFFAYIPKEOKYSFLONP 49  
|||||  
DB 27 FPTIPLSRFDNAMLRAHRLHQLAFDTYQFEFFAYIPKEOKYSFLONP 74

Search completed: September 15, 2003, 12:05:30  
Job time : 12.2401 secs

GenCore version 1.1.6  
Copyright (c) 1993 - 2003 Empugen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 12:03:35 ; Search time 20.5484 Seconds  
(without alignments)  
347.945 Million cell updates/sec

Title: US-09-423-100-1

Perfect score: 260

Sequence: 1 MF TIPLSRFDNMLRAHR.....OFFEEATIPKQYSFLQNP 49

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 641936 seqs, 14591426 residues

Total number of hits satisfying chosen parameters 641936

Minimum LR seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARY

Result No.	Score	Query Match	Length	DB ID	Description
1	260	100.0	49	14	US-10-054-873-1 Sequence 1, Appli
2	260	100.0	92	14	US-10-054-873-2 Sequence 2, Appli
3	260	100.0	107	14	US-10-054-873-6 Sequence 6, Appli
4	260	100.0	150	14	US-10-054-873-7 Sequence 7, Appli
5	255	98.1	191	11	US-09-984-010-23 Sequence 23, Appli
6	255	98.1	191	12	US-10-153-257-1 Sequence 1, Appli
7	255	98.1	191	12	US-10-400-77-1 Sequence 1, Appli
8	255	98.1	191	12	US-10-400-4-1 Sequence 1, Appli
9	255	98.1	214	12	US-10-153-7-6 Sequence 6, Appli
10	255	98.1	217	9	US-09-929-01-9 Sequence 9, Appli
11	255	98.1	245	9	US-09-280-01-66 Sequence 66, Appli
12	249	95.8	217	9	US-09-853-648-2 Sequence 2, Appli
13	249	95.8	217	12	US-09-969-48C-4 Sequence 4, Appli
14	248	95.4	217	10	US-09-804-409A-16 Sequence 16, Appli
15	242	93.1	217	9	US-09-853-688-4 Sequence 4, Appli

16	239	91.9	191	12	US-09-824-200-12	Sequence 12, Appli
17	199	76.5	217	9	US-09-850-887-3	Sequence 3, Appli
18	197	75.8	163	15	US-10-043-487-350	Sequence 350, App
19	197	75.8	191	12	US-10-153-207-2	Sequence 2, Appli
20	197	75.8	229	15	US-10-103-313-411	Sequence 411, App
21	188	72.3	246	15	US-10-188-246-18	Sequence 18, Appli
22	159.5	61.3	190	12	US-10-153-207-3	Sequence 3, Appli
23	142	54.6	54	9	US-09-876-478-14	Sequence 14, Appli
24	114	43.8	54	9	US-09-876-478-2	Sequence 2, Appli
25	114	43.8	54	9	US-09-876-478-3	Sequence 3, Appli
26	106	40.8	28	9	US-09-876-478-4	Sequence 4, Appli
27	106	40.8	28	9	US-09-876-478-7	Sequence 7, Appli
28	106	40.8	54	9	US-09-876-478-6	Sequence 6, Appli
29	88	33.6	41	15	US-10-191-879-22	Sequence 22, Appli
30	88	33.8	188	15	US-10-191-879-19	Sequence 19, Appli
31	88	33.8	210	15	US-10-191-879-10	Sequence 10, Appli
32	87	33.5	366	10	US-09-887-569A-2	Sequence 2, Appli
33	87	33.5	388	12	US-10-122-746-4	Sequence 4, Appli
34	83	31.9	56	9	US-09-876-478-15	Sequence 15, Appli
35	63	30.6	198	15	US-10-143-293-32	Sequence 32, Appli
36	60	30.6	199	15	US-10-143-293-33	Sequence 33, Appli
37	73	28.1	46	15	US-10-100-679-61	Sequence 61, Appli
38	73	28.1	46	15	US-10-100-679-62	Sequence 62, Appli
39	73	28.1	46	15	US-10-100-679-63	Sequence 63, Appli
40	72.5	27.9	125	14	US-10-036-869-25	Sequence 25, Appli
41	72.5	27.9	253	14	US-10-036-869-27	Sequence 27, Appli
42	71	27.3	197	15	US-10-140-293-13	Sequence 13, Appli
43	70	26.9	199	15	US-10-140-293-22	Sequence 22, Appli
44	70	26.9	199	15	US-10-140-293-23	Sequence 23, Appli
45	70	26.9	199	15	US-10-140-293-25	Sequence 25, Appli

#### ALIGNMENTS

RESULT 1  
US-10-054-873-1  
: Sequence 1, Application US/10054873  
: Publication No. US20020164712A1  
: GENERAL INFORMATION:  
: APPLICANT: Gan, Zhong Ru  
: TITLE OF INVENTION: Intramolecular Protein Containing an  
: NUMBER OF SEQUENCES: 7  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Townsend and Townsend and Crew LLP  
: STREET: Two Embarcadero Center, Eighth Floor  
: CITY: San Francisco  
: STATE: California  
: COUNTRY: USA  
: ZIP: 94111-3834  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/10/054,873  
: FILING DATE: 22-Jan-2002  
: CLASSIFICATION: <unknown>  
: PRIORITY APPLICATION DATA:  
: APPLICATION NUMBER: WO 98/0098/00052  
: FILING DATE: 31-MAR-1998  
: APPLICATION NUMBER: US 09/423,100  
: FILING DATE: 11-DEC-2000  
: ATTORNEY/AGENT INFORMATION:  
: NAME: MYCROFIT, Frank J  
: REGISTRATION NUMBER: 46,946  
: REFERENCE/DOCKET NUMBER: 020:67-000100US  
: INFORMATION FOR SEQ ID NO: 1:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 49 amino acids  
: TYPE: amino acid

```

; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-054-873-1

Query Match 100.0%; Score 260; DB 14; Length 49;
Best local Similarity 100.0%; Pred. No. 1 2e-28;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFPTPLSRFDNAMLKAHRLHQLAFDTYQFEFEAYIPKECKYSFLQNP 49
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
DB 1 MFPTPLSRFDNAMLKAHRLHQLAFDTYQFEFEAYIPKECKYSFLQNP 49

RESULT 2
US-10-054-873-2
; Sequence 2, Application US/10054873
; Publication No. US20020164712A;
; GENERAL INFORMATION:
; APPLICANT: Gan, Zheng Ku
; TITLE OF INVENTION: Chimeric Protein Containing an
; Intramolecular Chaperone-Like Sequence
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/054-873
; FILING DATE: 22-Jan-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/CN98 0952
; FILING DATE: 31-MAR-1998
; APPLICATION NUMBER: US 09/423,100
; FILING DATE: 11-DEC-2000
; ATTORNEY/AGENT INFORMATION:
; NAME: Mycroft, Frank J
; REGISTRATION NUMBER: 46,946
; REFERENCE/DOCKET NUMBER: 020167, 101300S
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 92 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-054-873-2

Query Match 100.0%; Score 260; DB 14; Length 92;
Best local Similarity 100.0%; Pred. No. 1 2e-27;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFPTPLSRFDNAMLKAHRLHQLAFDTYQFEFEAYIPKECKYSFLQNP 49
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
DB 1 MFPTPLSRFDNAMLKAHRLHQLAFDTYQFEFEAYIPKECKYSFLQNP 49

RESULT 3
US-10-054-873-6
; Sequence 6, Application US/10054873
; Publication No. US20020164712A;
; GENERAL INFORMATION:

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/CN98/00052
; FILING DATE: 31-MAR-1998
; APPLICATION NUMBER: US 09/423,100
; FILING DATE: 11-DEC-2000
; ATTORNEY/AGENT INFORMATION:
; NAME: Mycroft, Frank J
; REGISTRATION NUMBER: 46,946
; REFERENCE/DOCKET NUMBER: 020167-000130US
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 150 amino acids
; TYPE: amino acid
; STRANDEDNESS: <unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-10-054-873-7

Query Match 100.0%; Score 260; DB 14; Length 150;
Best Local Similarity 100.0%; Pred. No. 1.3e-26;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEPTPLSRFLDNAMLRHRLHQLAFDTYQFEFEAYIPKEQKYSFLONP 49
      |||||||
Db 1 MEPTPLSRFLDNAMLRHRLHQLAFDTYQFEFEAYIPKEQKYSFLONP 49

RESULT 5
US-09-984-010-23
; Sequence 23, Application US/09984010
; Publication No. US20030104578A1
; GENERAL INFORMATION:
; APPLICANT: Ballance, David James
; TITLE OF INVENTION: RECOMBINANT FUSION PROTEINS TO GROWTH HORMONE
; AND SERUM ALBUMIN
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNEK, LLP
; STREET: 1300 I Street, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.70 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/984,010
; FILING DATE: 21-May-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/091,873
; FILING DATE: 25-JUN-1998
; APPLICATION NUMBER: PCT/CN96/03164
; FILING DATE: 19-DEC-1996
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; STRANDEDNESS: <unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-09-984-010-23

Query Match 98.1%; Score 255; DB 11; Length 191;
Best Local Similarity 100.0%; Pred. No. 1.3e-26;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/CN98/00052
; FILING DATE: 31-MAR-1998
; APPLICATION NUMBER: US 09/423,100
; FILING DATE: 11-DEC-2000
; ATTORNEY/AGENT INFORMATION:
; NAME: Mycroft, Frank J
; REGISTRATION NUMBER: 46,946
; REFERENCE/DOCKET NUMBER: 020167-000130US
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 150 amino acids
; TYPE: amino acid
; STRANDEDNESS: <unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-10-054-873-7

Query Match 100.0%; Score 260; DB 14; Length 150;
Best Local Similarity 100.0%; Pred. No. 1.3e-26;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEPTPLSRFLDNAMLRHRLHQLAFDTYQFEFEAYIPKEQKYSFLONP 49
      |||||||
Db 1 MEPTPLSRFLDNAMLRHRLHQLAFDTYQFEFEAYIPKEQKYSFLONP 49

RESULT 6
US-10-153-207-1
; Sequence 1, Application US/10153207
; Publication No. US20030153003A1
; GENERAL INFORMATION:
; APPLICANT: James A. Wells
; APPLICANT: Brian C. Cunningham
; TITLE OF INVENTION: GROWTH HORMONE VARIANTS
; FILE REFERENCE: 669.12-US-C7
; CURRENT APPLICATION NUMBER: US/10/153,207
; CURRENT FILING DATE: 2002-05-22
; PRIOR APPLICATION NUMBER: 08/479,884
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 08/199,723
; PRIOR FILING DATE: 1994-02-02
; PRIOR APPLICATION NUMBER: 07/969,227
; PRIOR FILING DATE: 1992-10-13
; PRIOR APPLICATION NUMBER: 07/875,204
; PRIOR FILING DATE: 1992-04-27
; PRIOR APPLICATION NUMBER: 07/428,066
; PRIOR FILING DATE: 1989-10-26
; PRIOR APPLICATION NUMBER: 07/264,611
; PRIOR FILING DATE: 1988-10-28
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-153-207-1

Query Match 98.1%; Score 255; DB 12; Length 191;
Best Local Similarity 100.0%; Pred. No. 1.3e-26;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PPTPLSRFLDNAMLRHRLHQLAFDTYQFEFEAYIPKEQKYSFLONP 49
      |||||||
Db 1 PPTPLSRFLDNAMLRHRLHQLAFDTYQFEFEAYIPKEQKYSFLONP 49

RESULT 7
US-10-400-377-1
; Sequence 1, Application US/10400377
; Publication No. US20030162949A1
; GENERAL INFORMATION:
; APPLICANT: Cox III, George N
; APPLICANT: Bolder Biotechnology, Inc.
; TITLE OF INVENTION: Derivatives of Growth Hormone and Related Proteins
; FILE REFERENCE: 4152-1-PUS
; CURRENT APPLICATION NUMBER: US/10/400,377
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US/09/462,941
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: 60/052,516
; PRIOR FILING DATE: 1997-07-14
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-400-377-1

Query Match 98.1%; Score 255; DB 12; Length 191;
Best Local Similarity 100.0%; Pred. No. 1.3e-26;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PPTPLSRFLDNAMLRHRLHQLAFDTYQFEFEAYIPKEQKYSFLONP 49
      |||||||
Db 1 PPTPLSRFLDNAMLRHRLHQLAFDTYQFEFEAYIPKEQKYSFLONP 49
```

Db 1 FTIPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEOKYSFLQNP 48  
|||||  
RESULT 8  
US-10-400-708-1  
; Sequence 1, Application US/10400708  
; Publication No. US2003016865A1  
; GENERAL INFORMATION:  
; APPLICANT: Cox III, George N  
; TITLE OF INVENTION: Derivatives of Growth Hormone and Related Proteins  
; FILE REFERENCE: 4152-1-PUS  
; CURRENT APPLICATION NUMBER: US/10/400,708  
; CURRENT FILING DATE: 2003-03-26  
; PRIOR APPLICATION NUMBER: US/09/462,941  
; PRIOR FILING DATE: 2000-01-14  
; PRIOR APPLICATION NUMBER: 60/052,516  
; PRIOR FILING DATE: 1997-07-14  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 191  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-400-708-1

Query Match 98.1%; Score 255; DB 12; Length 191;  
Best Local Similarity 100.0%; Pred. No. 1.5e-26;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 FTIPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEOKYSFLQNP 49  
Db 1 FTIPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEOKYSFLQNP 48  
|||||

RESULT 9  
US-10-151-207-6  
; Sequence 6, Application US/10153207  
; Publication No. US20030153003A1  
; GENERAL INFORMATION:  
; APPLICANT: James A. Wells  
; APPLICANT: Brian C. Cunningham  
; TITLE OF INVENTION: GROWTH HORMONE VARIANT  
; FILE REFERENCE: 669-12 US-67  
; CURRENT APPLICATION NUMBER: US/10/153,207  
; CURRENT FILING DATE: 2002-05-22  
; PRIOR APPLICATION NUMBER: 08/479,884  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 08/190,723  
; PRIOR FILING DATE: 1994-02-02  
; PRIOR APPLICATION NUMBER: 07/960,227  
; PRIOR FILING DATE: 1992-10-11  
; PRIOR APPLICATION NUMBER: 07/875,204  
; PRIOR FILING DATE: 1992-04-27  
; PRIOR APPLICATION NUMBER: 07/428,056  
; PRIOR FILING DATE: 1989-10-26  
; PRIOR APPLICATION NUMBER: 07/264,611  
; PRIOR FILING DATE: 1988-10-28  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 214  
; TYPE: PRT  
; ORGANISM: Homo Sapiens  
US-10-151-207-6

Query Match 98.1%; Score 255; DB 12; Length 214;  
Best Local Similarity 100.0%; Pred. No. 1.5e-26;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 FTIPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEOKYSFLQNP 49  
|||||

Db 24 FTIPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEOKYSFLQNP 71  
|||||  
RESULT 10  
US-09-929-918-9  
; Sequence 9, Application US/09929918  
; Patent No. US20020090678A1  
; GENERAL INFORMATION:  
; APPLICANT: Kordyum, Vitaliy A.  
; APPLICANT: Cherykh, Svitlana I.  
; APPLICANT: Slavchenko, Iryna Yu.  
; APPLICANT: Vozianov, Oleksandr  
; TITLE OF INVENTION: PHAGE-DEPENDENT SUPER PRODUCTION OF  
; FILE REFERENCE: BIOLOGICALLY ACTIVE PROTEIN AND PEPTIDES  
; CURRENT APPLICATION NUMBER: US/09/929,918  
; CURRENT FILING DATE: 2001-08-15  
; PRIOR APPLICATION NUMBER: 09/314,288  
; PRIOR FILING DATE: 1999-05-25  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9  
; LENGTH: 217  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-929-918-9

Query Match 98.1%; Score 255; DB 9; Length 217;  
Best Local Similarity 100.0%; Pred. No. 1.5e-26;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 FTIPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEOKYSFLQNP 49  
Db 27 FTIPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEOKYSFLQNP 74  
|||||

RESULT 11  
US-09-280-030-66  
; Sequence 66, Application US/092803030A  
; Patent No. US2001002155A1  
; GENERAL INFORMATION:  
; APPLICANT: Sato, Seiji  
; APPLICANT: Higashikuni, Naohiko  
; APPLICANT: Kudo, Toshiyuki  
; APPLICANT: Kondo, Masaaki  
; TITLE OF INVENTION: DNAS ENCODING NEW FUSION PROTEINS AND PROCESSES FOR THE  
; TITLE OF INVENTION: PREPARING USEFUL POLYPEPTIDES THROUGH EXPRESSION OF THE  
; FILE REFERENCE: 382,1026  
; CURRENT APPLICATION NUMBER: US/09/280,030A  
; CURRENT FILING DATE: 1999-03-26  
; EARLIER APPLICATION NUMBER: JP10-87335/1998  
; EARLIER FILING DATE: 1998-03-31  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 66  
; LENGTH: 245  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Designated is  
; OTHER INFORMATION: an amino acid sequence of MWpsp-MWp20-TEV-G-CH  
US-09-280-030-66

Query Match 98.1%; Score 255; DB 9; Length 245;  
Best Local Similarity 100.0%; Pred. No. 1.7e-26;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 FTIPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEOKYSFLQNP 49  
Db 55 FTIPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEOKYSFLQNP 102  
|||||



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:56:45 ; Search time 8.07885 Seconds  
(without alignments)  
583.284 Million cell updates/sec

Title: US-09-423-100-1

Perfect score: 260

Sequence: 1 MFPTIPLSRLFDNAHLRAH.....QEFEEAIPKQKYSFLQNP 49

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 46168642 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 25000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARY

Result No.	Score	Query Match	Length	ID	Description
1	255	98.1	217	1	STHU
2	255	98.1	217	2	somatotropin 1 pre
3	228	87.7	217	1	STHUV
4	228	87.7	256	1	somatotropin 2 pre
5	213	81.9	212	2	somatotropin 2 pre
6	213	81.9	212	2	chorionic somatoma
7	205	78.8	217	2	chorionic somatoma
8	201	77.4	217	2	somatotropin - the
9	197	75.8	215	2	chorionic somatoma
10	197	75.8	217	1	chorionamniotropon
11	197	75.8	217	1	chorionamniotropon
12	161.5	62.1	216	1	STWS
13	160.5	61.7	190	2	PN0140
14	159.5	61.3	190	1	STHO
15	159.5	61.3	190	2	JK0219
16	159.5	61.3	190	2	JK0429
17	159.5	61.3	216	1	STRT
18	159.5	61.3	216	1	STPG
19	159.5	61.3	216	2	146145
20	159.5	61.3	216	2	S49483
21	159.5	61.3	216	2	S49159
22	159.5	61.3	216	2	JK4632
23	156.5	60.2	216	2	A37782
24	155.5	59.8	190	1	A61584
25	150	57.7	216	2	JK1514
26	148	56.9	191	2	A60625
27	146	56.2	163	2	JK0387
28	144	55.4	190	2	S21750
29	144	55.4	216	2	A60509

30	142.5	54.8	217	1	STBO
31	142.5	54.8	217	1	STSH
32	142.5	54.8	217	1	STGT
33	142.5	54.8	217	2	S32682
34	140	53.8	216	2	S04929
35	132	50.8	190	2	A56816
36	132	50.8	215	2	I51188
37	128	49.2	195	2	I51250
38	128	49.2	215	2	JS0037
39	122	46.4	199	2	B32435
40	116	44.6	183	2	A60623
41	98.5	37.9	87	4	I67761
42	97	37.3	200	2	I51114
43	87	33.5	210	2	S69263
44	87	33.5	210	2	S69262
45	87	33.5	210	2	S02764

## ALIGNMENTS

### RESULT 1

STHU

somatotropin 1 precursor [validated] - human  
N:Alternate names: growth hormone 1; hGH-N; pituitary somatotropin  
N:Contains: growth hormone 5K peptide; somatotropin 1, long form; somatotropin 1, st  
C:Species: Homo sapiens (man)  
C:Date: 24-Apr-1984 #sequence-revision 10-Feb 1995 #text-change 08-Dec-2000  
C:Accession: A93731; A32435; A93694; A94247; A90051; A93397; A93778; A91764; A90217;  
R:DeMoto, F.M.; Moore, D.D.; Goodman, H.M.  
Nucleic Acids Res. 9, 3719-3730, 1981  
A:Title: Human growth hormone DNA sequence and mRNA structure: possible alternative  
A:Reference number: A93731; MUIB:82014939; PMID:6249091  
A:Accession: A93731;  
A:Molecule type: DNA  
A:Residues: 1-217 <DEN>  
A:Cross-references: GB:V00520  
R:Chen, E.Y.; Liao, Y.C.; Smith, D.H.; Barrera-Saldana, H.A.; Gellinas, R.E.; Seebur-  
Genomics 4, 479-497, 1989  
A:Title: The human growth hormone locus: nucleotide sequence, biology, and evolution  
A:Reference number: A32435; MUID:83307277; PMID:2744760  
A:Accession: A32435  
A:Molecule type: DNA  
A:Residues: 1-217 <CHP>  
A:Cross-references: GB:303071; NID:q183148; PIDN:AAA52549.1; PID:q183149  
R:Roskam, W.; Rougeon, F.  
Nucleic Acids Res. 7, 305-320, 1979  
A:Title: Molecular cloning and nucleotide sequence of the human growth hormone struc-  
A:Reference number: A93694; MUIB:80034477; PMID:386281  
A:Accession: A93694  
A:Molecule type: mRNA  
A:Residues: 1-217 <RGS>  
A:Cross-references: GB:V00519  
R:Mattial, J.A.; Halliwell, R.A.; Baxter, J.D.; Goodman, H.M.  
Science 205, 602-607, 1979  
A:Title: Human growth hormone: complementary DNA cloning and expression in bacteria.  
A:Reference number: A94247; MUID:79203293; PMID:377496  
A:Accession: A94247  
A:Molecule type: mRNA  
A:Residues: 1-217 <MAR>  
R:Li, C.H.; Dixon, J.S.; Liu, W.K.  
Arch. Biochem. Biophys. 133, 70-91, 1969  
A:Title: Human pituitary growth hormone. XIX. The primary structure of the hormone.  
A:Reference number: A90048; MUID:69289202; PMID:5810834  
A:Contents: annotation  
R:Li, C.H.; Dixon, J.S.  
Arch. Biochem. Biophys. 146, 233-236, 1971  
A:Title: Human pituitary growth hormone. XXXII. The primary structure of the hormone.  
A:Reference number: A90051; MUID:72143935; PMID:5144027  
A:Accession: A90051  
A:Molecule type: protein

A:Residues: 27-94;96-217 <LIC>  
 R:Niall, H.D.  
 Nature New Biol. 230, 90-91, 1971  
 A:Title: Revised primary structure for human growth hormone.  
 A:Reference number: A93397; MUID:71139765; PMID:5279046  
 A:Accession: A93397  
 A:Molecule type: protein  
 A:Residues: 27-51 <NIA>  
 R:Niall, H.D.; Hogan, M.L.; Sauer, R.; Rosenblum, I.V.; Greenwood, F.C.  
 Proc. Natl. Acad. Sci. U.S.A. 68, 866-869, 1971  
 A:Title: Sequences of pituitary and placental lactogenic and growth hormones: evolution  
 A:Reference number: A93778; MUID:71153968; PMID:5279528  
 A:Accession: A93778  
 A:Molecule type: protein  
 A:Residues: 119-120;157-159 <NI2>  
 R:Niall, H.D.  
 in Prolactin and Carcinogenesis, Proc. Fourth Tenovus Workshop Prolactin, Griffiths, K.,  
 A:Title: The Chemistry of the human lactogenic hormones.  
 A:Reference number: A94427  
 R:Contents: annotation; somatotropin revision  
 R:Bewley, T.A.; Jirik, J.S.; Liu, C.H.  
 Int. J. Pept. Protein Res. 4, 281-287, 1972  
 A:Title: Sequence comparison of human pituitary growth hormone, human chorionic somatomedin  
 A:Reference number: A91764; MUID:73022628; PMID:477544  
 A:Accession: A91764  
 A:Molecule type: protein  
 A:Residues: 27-217 <REW>  
 R:Lewis, U.J.; Bonawald, L.F.; Lewis, L.J.  
 Biochem. Biophys. Res. Commun. 92, 511-516, 1980  
 A:Title: The 20,000-dalton variant of human growth hormone: location of the amino acid  
 A:Reference number: A90217; MUID:80130196; PMID:7356479  
 R:Contents: somatotropin, 20K short variant  
 A:Accession: A90217  
 A:Molecule type: protein  
 A:Residues: 45, 57-74-80 <LEW>  
 R:Chapman, G.E.; Rogers, K.M.; Brittain, T.; Bradshaw, R.A.; Bates, G.J.; Turner, C.; Ca  
 J. Biol. Chem. 256, 2395-2401, 1981  
 A:Title: The 20,000 molecular weight variant of human growth hormone. Preparation and so  
 A:Reference number: A92311; MUID:81117361; PMID:7462247  
 R:Contents: somatotropin, 20K short variant  
 A:Accession: A92311  
 A:Molecule type: protein  
 A:Residues: 27-57;73-79 <CHA>  
 R:Singh, R.N.P.; Seavey, B.K.; Lewis, L.J.; Lewis, H.J.  
 J. Protein Chem. 2, 425-436, 1983  
 A:Title: Human growth hormone peptide 1-43: isolation from pituitary glands.  
 A:Reference number: A61466  
 A:Accession: A61466  
 A:Molecule type: protein  
 A:Residues: 27-69 <SIN>  
 R:Robson, V.M.J.; Rae, I.D.; NG, F.  
 Biol. Chem. Hoppe-Seyler 371, 423-431, 1990  
 A:Title: Identification of the aspartamide structure in a previously-reported peptide.  
 A:Reference number: S09685; MUID:90334745; PMID:2378679  
 A:Accession: S09685  
 A:Molecule type: protein  
 A:Residues: 27-14;17,16-47 <ROB>  
 R:De Vos, A.M.; Uitsch, M.; Kossiakoff, A.A.  
 Science 255, 306-312, 1992  
 A:Title: Human growth hormone and extracellular domain of its receptor: crystal structu  
 A:Reference number: A41728; MUID:92196577; PMID:1549776  
 R:Contents: annotation; X-ray crystallography, 2.8 angstroms  
 A:Note: The structure of the complex with growth hormone receptor is described  
 R:Gray, G.L.; Baldrige, J.S.; McKeown, K.S.; Heyneker, H.L.; Chang, C.N.  
 Gene 19, 247-254, 1985  
 A:Title: Periplasmic production of correctly processed human growth hormone in Escherich  
 A:Reference number: I41126; MUID:86137393; PMID:3912261  
 A:Accession: I84549  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-26 <RES>  
 A:Cross-references: GB:M14398; NID:g183158; PIDN:AAA52554.1; PID:g183159

C:Comment: The gene for this hormone is transcribed only in somatotrophic cells of th  
 C:Comment: About 90% of somatotropin is the 22K long form.  
 C:Genetics:  
 A:Gene: GDB:GH1  
 A:Cross-references: GDB:119982; OMIM:139250  
 A:Map position: 17q23.1-17q23.3  
 A:Introns: 4/1; 57/3; 97/3; 152/3  
 C:Superfamily: prolactin  
 C:Keywords: alternative splicing; hormone; pituitary  
 F:1-26/Domain: signal sequence #status predicted <Sig>  
 F:27-217/Product: somatotropin 1, long form #status experimental <SOI>  
 F:27-69/Product: growth hormone 5K peptide #status experimental <SKP>  
 F:27-57,73-217/Product: somatotropin 1, short form #status experimental <SOS>  
 F:79-191,208-215/Disulfide bonds: #status experimental

Query Match 98.1%; Score 255; DB 1; Length 217;  
 Best Local Similarity 100.0%; Pred. No. 2e-24;  
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 FVTIFSLFNNAMLRKHLHQLAFDTYQFFFEAYIPKEQKYSFLQNP 49  
 DB 27 FVTIFSLFNNAMLRKHLHQLAFDTYQFFFEAYIPKEQKYSFLQNP 74

RESULT 2  
 167410  
 somatotropin - rhesus macaque  
 N:Alternate names: growth hormone  
 C:Species: Macaca mulatta (rhesus macaque)  
 C:Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 16-Jul-1999  
 C:Accession: I67410; A05094  
 R:Gollos, T.G.; Burning, M.; Fisher, J.M.; Fowler, P.D.  
 Endocrinology 133, 1744-1752, 1993  
 A:Title: Cloning of four growth hormone/chorionic somatomammotropin-related complene  
 A:Reference number: I53267; MUID:94008724; PMID:8414617  
 A:Accession: I67410  
 A:Status: translated from GB/EMBL/DDBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-217 <RES>  
 A:Cross-references: GB:IJ6556; NID:g291114; PIDN:AAA18842.1; PID:g293115  
 R:Li, C.H.; Chung, D.; Lahm, H.W.; Stein, S.  
 Arch. Biochem. Biophys. 245, 287-291, 1986  
 A:Title: The primary structure of monkey pituitary growth hormone.  
 A:Reference number: A05094; MUID:86129460; PMID:3080959  
 A:Accession: A05094  
 A:Molecule type: protein  
 A:Residues: 27-99; Q'101-178; D'180-217 <IIR>  
 A:Note: The monkey species is not identified in the reference  
 R:Raben, M.S.  
 Science 125, 883-884, 1957  
 A:Title: Preparation of growth hormone from pituitaries of man and monkey.  
 A:Reference number: A44774  
 R:Contents: annotation; Identification of source organism  
 C:Superfamily: prolactin

Query Match 98.1%; Score 255; DB 2; Length 217;  
 Best Local Similarity 100.0%; Pred. No. 2e-24;  
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 FVTIFSLFNNAMLRKHLHQLAFDTYQFFFEAYIPKEQKYSFLQNP 49  
 DB 27 FVTIFSLFNNAMLRKHLHQLAFDTYQFFFEAYIPKEQKYSFLQNP 74

RESULT 3  
 STHUV  
 somatotropin 2 precursor - human  
 N:Alternate names: growth hormone 2; growth hormone variant; hGH-V; placental somatot  
 C:Contains: somatotropin 2, long splice form; somatotropin 2, short splice form  
 C:Species: Homo sapiens (man)  
 C:Date: 17-Dec-1982 #sequence\_revision 10-Feb-1995 #text\_change 21-Jul-2000  
 C:Accession: D32435; R28072; A01511; I52104; A6071;  
 R:Chen, F.Y.; Liao, Y.C.; Smith, D.H.; Barrera-Saldana, H.A.; Gelinas, R.E.; Seeburg

Genomics 4. 479-497, 1999

A:Title: The human growth hormone locus: nucleotide sequence, biology, and evolution.  
A:Reference number: A32435; MUID:89307277; PMID:2744760  
A:Accession: D32435  
A:Molecule type: DNA  
A:Residues: 1-217 <CHE>  
A:Cross-references: GB:J03071; NID:g183148; PIDN:AA52552.1; PID:g183152  
R:Cooke, N.E.; Ray, J.; Emery, J.G.; Liebhabor, S.A.  
J. Biol. Chem. 263, 9001-9006, 1988  
A:Title: Two distinct species of human growth hormone-variant mRNA in the human placenta  
A:Reference number: A92725; MUID:88243769; PMID:3379057  
A:Accession: A28072  
A:Molecule type: mRNA  
A:Residues: 1-217 <COO>  
R:Seeburg, P.H.  
DNA 1, 239-249, 1982  
A:Title: The human growth hormone gene family: nucleotide sequences show recent divergence  
A:Reference number: A31511; MUID:83352010; PMID:2169099  
A:Accession: A31511  
A:Molecule type: DNA  
A:Residues: 1-34, 136-217 <SH>  
R:Igout, A.; Scipio, M.D.; Frankenne, F.; Beuvenen, J.; Igout, A.; Henner, G.  
Arch. Int. Physiol. Biochim. 96, 63-67, 1988  
A:Title: Cloning and nucleotide sequence of placental hGH-V cDNA.  
A:Reference number: 152104; MUID:8904984; PMID:2460050  
A:Accession: F52104  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-217 <ICO>  
A:Cross-references: GB:M38451; NID:g183179; PIDN:AAA35891.1; PID:g183180  
R:Frankenne, F.; Scipio, M.L.; Van Beuvenen, J.; Igout, A.; Henner, G.  
J. Clin. Endocrinol. Metab. 71, 15-18, 1990  
A:Title: Identification of placental human growth hormone as the growth hormone-V gene  
A:Reference number: A60711; MUID:50317018; PMID:2196278  
A:Accession: A60711  
A:Molecule type: protein  
A:Residues: 27-44; 46-57 <PRA>  
A:Experimental source: tissue placenta  
A:Note: partial glycosylation was demonstrated by lectin binding  
C:Comment: This gene is expressed by the placenta.  
C:Genetics:  
A:Gene: GDB:GH2  
A:Cross references: GDB:119983; OMIM:149240  
A:Map position: 17q22-17q24  
A:Introns: 4/1; 57/3; 97/3; 152/3  
C:Superfamily: prolactin  
C:Keywords: alternative splicing; glycoprotein; hormone; placenta  
F:1-26/domain; signal sequence #status predicted <SIG>  
F:27-217/product; somatotropin 2, long splice form #status predicted <SL>  
F:27-57, 73-217/product; somatotropin 2, short splice form #status predicted <SOS>  
F:79-191, 208-215/disulfide bonds; #status predicted  
F:166/binding site; carbohydrate (Asn) (covalent) #status predicted  
Query Match 87.7%; Score 228; DB 1; Length 217;  
Best local similarity 91.7%; Pred. No. 4 60-21;  
Matches 44; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 FETIPLSRFLDNAMLRHRIHQIAFTGYOEFEPAYIPKQKYSFLQNP 49  
|||||  
DB 27 FETIPLSRFLDNAMLRRLYQIAYDTGYOEFEPAYI:KEQKYSFLQNP 74  
|||||

RESULT 4  
STHUV2  
SOMATOTROPIN 2 precursor, splice form 2 - human  
N:Alternate names: growth hormone variant-2; placental somatotropin form 2  
C:Species: Homo sapiens (man)  
C:Date: 10-Sep-1989 #sequence\_revision 10-Feb-1995 #text\_change 02-Sep-1997  
C:Accession: A28072  
R:Cooke, N.E.; Ray, J.; Emery, J.G.; Liebhabor, S.A.  
J. Biol. Chem. 263, 9001-9006, 1988  
A:Title: Two distinct species of human growth hormone-variant mRNA in the human placenta  
A:Reference number: A92725; MUID:88243769; PMID:3379057



[illegible]



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:54:00 ; Search time 4.56631 seconds  
(without alignments)  
504.633 Million cell updates/sec

Title: US-09-423-100-1

Perfect score: 260

Sequence: 1 MPPTIPLSLFDNAMI,RAHR.....QEFPEAYTPKEQKYSFLQNP 49

Scoring table: HXSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127864 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127864

Minimum hit seq length: 6

Maximum hit seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	255	98.1	217	1	SOMA_HUMAN
2	255	98.1	217	1	SOMA_MACMU
3	255	98.1	217	1	SOMA_PANTR
4	249	95.8	217	1	SOMA_CALJA
5	249	95.8	217	1	SOMA_SALIB
6	236	90.8	217	1	SOMA_PANTR
7	228	87.7	217	1	SOMA2_HUMAN
8	199	76.5	217	1	SOMA2_MACMU
9	197	75.8	217	1	PIL_HUMAN
10	161.5	62.1	216	1	SOMA_MUISE
11	160.5	61.7	190	1	SOMA_BALWO
12	153.5	61.3	190	1	SOMA_LOXAF
13	159.5	61.3	190	1	SOMA_VULVU
14	159.5	61.3	216	1	SOMA_CANFA
15	159.5	61.3	216	1	SOMA_FELCA
16	159.5	61.3	216	1	SOMA_HORSE
17	159.5	61.3	216	1	SOMA_MESAU
18	159.5	61.3	216	1	SOMA_PIG
19	159.5	61.3	216	1	SOMA_RAHIT
20	159.5	61.3	216	1	SOMA_RAT
21	159.5	61.3	217	1	SOMA_CALSF
22	159.5	61.3	217	1	SOMA_NYCPY
23	156.5	60.2	216	1	SOMA_MUSVI
24	155.5	59.8	190	1	SOMA_LAMIA
25	150	57.7	216	1	SOMA_MELGA
26	148	56.9	191	1	SOMA_CHEMY
27	145	55.8	215	1	SOMA_MONDO
28	145	55.8	215	1	SOMA_TRIVU
29	144	55.4	190	1	SOMA1_ACIGU
30	144	55.4	190	1	SOMA2_ACIGU
31	144	55.4	216	1	SOMA_CHICK
32	142.5	54.8	217	1	SOMA_BOVIN
33	142.5	54.8	217	1	SOMA_CEREL

34 142.5 54.8 217 1 SOMA\_SHEEP  
35 142 54.6 217 1 SOMA\_STRCA  
36 140 53.8 190 1 SOMA\_CROKO  
37 140 53.8 216 1 SOMA\_ANAPL  
38 135.5 52.1 217 1 SOMA\_BUBBU  
39 132 50.8 215 1 SOMA\_RANCA  
40 125 48.1 211 1 SOMA\_LEPOS  
41 122 46.9 214 1 SOMA\_XENLA  
42 116 44.6 163 1 SOMA\_PIGL  
43 112 43.1 206 1 SOMA\_PROAN  
44 111 42.7 213 1 SOMA\_HUFMA  
45 104 40.0 208 1 SOMB\_XENLA

PC1247 ovis aries  
Q9PW93 struthio ca  
P55755 crocodylus  
P11228 anas platyr  
O18938 bubalus bub  
P10813 rana catesb  
P79885 lepisosteus  
P12855 xenopus lae  
P34006 prionace gl  
O73848 protoporus  
O73849 bufo marinus  
P12856 xenopus lae

#### ALIGNMENTS

RESULT 1  
SOMA\_HUMAN  
ID SOMA\_HUMAN STANDARD; PRT: 217 AA.  
AC P01241: Q14495, Q16641, Q9H871, Q9CMU7, Q9UNI5;  
DI 21-JUL-1986 (Ref: 01, Created)  
DI 01-MAR-1992 (Ref: 21, Last sequence update)  
DI 15-SEP-2003 (Ref: 42, Last annotation update)  
DE Somatotropin precursor (growth hormone) (GH) (GH N) (Pituitary growth hormone) (Growth hormone 1)  
GN GH1  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chorda i; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID:9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE=80034477; PubMed=386281;  
RA Roskam W., Rougeon F.  
RT \*Molecular cloning and nucleotide sequence of the human growth hormone structural gene.\*  
RT Nucleic Acids Res. 7:305-320(1979).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE=79203293; PubMed=377496;  
RA Martini J.A., Hallelwell R.A., Baxter J.D., Goodman H.M.;  
RT \*Human growth hormone: complementary DNA cloning and expression in bacteria.\*  
RT Science 205:602-607(1979).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 1), AND POSSIBLE ALTERNATIVE SPLICING.  
RX MEDLINE=82014939; PubMed=6269091;  
RA Denoto F.M., Moore D.D., Goodman H.M.;  
RT \*Human growth hormone DNA sequence and mRNA structure: possible alternative splicing.\*  
RT Nucleic Acids Res. 9:3719-3730(1981).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=83182010; PubMed=7169009;  
RA Seeburg P.H.;  
RT \*The human growth hormone gene family: nucleotide sequences show recent divergence and predict a new polypeptide hormone.\*;  
RT DNA 1:239-249(1982).  
RN [5]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89307277; PubMed=2744760;  
RA Chen E.Y., Liao Y.C., Smith D.H., Herrera-Saldana H.A.,  
RT Gelinas R.E., Seeburg P.H.;  
RT \*The human growth hormone locus: nucleotide sequence, biology, and evolution.\*;  
RT Genomics 4:479-497(1989).  
RN [6]  
RP SEQUENCE FROM N.A. (ISOFORM 3).  
RC TISSUE=Pituitary;  
RA Gu J., Huang Q.-H., Li N., Xu S.-H., Han Z.-G., Fu G., Chen Z.;  
RT \*A novel gene expressed in human pituitary.\*;  
RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A. (ISOFORM 4).  
 RC TISSUE-PITUITARY; PubMed:10931946;  
 RX MEDLINE-20402571; PubMed:10931946;  
 RA Hu R.-M., Han Z.-G., Song H.-D., Peng Y. D., Huang Q.-H., Ren S.-X.,  
 RA Gu Y.-J., Huang C.-H., Li Y.-B., Jiang C. L., Fu G., Zhang Q.-H.,  
 RA Gu B.-W., Dai M., Mao Y.-F., Gao G.-F., Xiong R., Ye M., Zhou J.,  
 RA Xu S.-H., Gu J., Shi J.-X., Jin W.-R., Zhang C.-K., Wu T.-M.,  
 RA Huang G.-Y., Chen Z., Chen M.-D., Chen J.-L.,  
 RT "Gene expression profiling in the human hypothalamus-pituitary-adrenal  
 axis and full-length cDNA cloning.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:9543-9548(2000).  
 RN [17]  
 RP SEQUENCE OF 1-26 FROM N.A.  
 RX MEDLINE-86137493; PubMed:3912261;  
 RA Gray G.L., Baldridge J.S., McKewen K.S., Heyneker H.L., Chang C.N.,  
 RT "Periplasmic production of correctly processed human growth hormone in  
 Escherichia coli: natural and bacterial signal sequences are  
 interchangeable.";  
 RL Gene 99:247-254(1985).  
 RN [19]  
 RP SEQUENCE OF 27-217.  
 RX MEDLINE-69289402; PubMed 5810834;  
 RA Li C.H., Dixon J.S., Liu W.-K.,  
 RT "Human pituitary growth hormone. XIX. The primary structure of the  
 hormone.";  
 RL Arch. Biochem. Biophys. 133:70-91(1965).  
 RN [10]  
 RP SEQUENCE OF 27-217, AND REVISIONS.  
 RX MEDLINE-72143935; PubMed:5144027;  
 RA Li C.H., Dixon J.S.,  
 RT "Human pituitary growth hormone. 32. Th. primary structure of the  
 hormone: revision.";  
 RL Arch. Biochem. Biophys. 146:233-236(1971).  
 RN [11]  
 RP REVISION.  
 RX MEDLINE-73092028; PubMed:4675454;  
 RA Bewley T.A., Dixon J.S., Li C.H.,  
 RT "Sequence comparison of human pituitary growth hormone, human  
 chorionic somatomotropin, and ovine pituitary growth and  
 lactogenic hormones.";  
 RL Int. J. Pept. Protein Res. 4:281-287(1972).  
 RN [12]  
 RP SEQUENCE OF 27-61 AND 102-124.  
 RX MEDLINE-71189765; PubMed:5279046;  
 RA Niall H.D.,  
 RT "Revised primary structure for human growth hormone.";  
 RL Nature New Biol. 230:90-91(1971).  
 RN [13]  
 RP REVISIONS TO 119-120 AND 157-159.  
 RX MEDLINE 71154968; PubMed:5279528;  
 RA Niall H.D., Hoan M.L., Sauer R., Rosenblum I.Y., Greenwood F.C.,  
 RT "Sequences of pituitary and placental lactogenic and growth hormones:  
 evolution from a primordial peptide by gene reduplication.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 68:866-869(1971).  
 RN [14]  
 RP REVISION.  
 RA Niall H.D.,  
 RT "The chemistry of the human lactogenic hormones.";  
 RL (In) Griffiths R. (eds.),  
 RL Prolactin and carcinogenesis. Proc. fourth renovus workshop prolactin,  
 RL pp.14-20. Alpha Omega Alpha Press, Cardiff (1972).  
 RN [15]  
 RP SEQUENCE OF 27-79 (ISOFORM 2).  
 RX MEDLINE-81117461; PubMed:7462247;  
 RA Chapman G.E., Rogers K.M., Brittain T., Bradshaw R.A., Bates G.J.,  
 RA Turner C., Cary P.D., Crane-Robinson C.,  
 RT "The 20,000 molecular weight variant of human growth hormone.  
 Preparation and some physical and chemical properties.";  
 RL J. Biol. Chem. 256:2395-2401(1981).  
 RN [16]  
 RP SEQUENCE OF 46-80 (ISOFORM 2).  
 RX MEDLINE-80130196; PubMed:7356479;  
 Lewis U.J., Bonewald L.F., Lewis L.J.,  
 RT "The 20,000-dalton variant of human growth hormone: location of the  
 amino acid deletions.";  
 RL Biochem. Biophys. Res. Commun. 92:511-516(1980).  
 RN [17]  
 RP DEMILCATION OF GLN-163 AND ASN-178.  
 RX MEDLINE-82052997; PubMed:7028740;  
 RA Lewis U.J., Singh R.N., Bonewald L.F., Seavey B.K.,  
 RT "Altered proteolytic cleavage of human growth hormone as a result of  
 deamidation.";  
 RL J. Biol. Chem. 256:11645-11650(1981).  
 RN [18]  
 RP REVIEW.  
 RX MEDLINE-99321812; PubMed:10393484;  
 RA Baumann G.,  
 RT "Growth hormone heterogeneity in human pituitary and plasma.";  
 RL Horm. Res. 51 Suppl. 1:2-6(1999).  
 RN [19]  
 RP 3D-STRUCTURE MODELING.  
 RX MEDLINE 88196073; PubMed 444173;  
 RA Cohen P.E., Kuntz L.D.,  
 RT "Prediction of the three dimensional structure of human growth  
 hormone.";  
 RL Proteins 2:162-156(1987).  
 RN [20]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).  
 RX MEDLINE-92196577; PubMed:1549776;  
 RA de Vos A.M., Ullsch M., Kossiakoff A.A.,  
 RT "Human growth hormone and extracellular domain of its receptor:  
 crystal structure of the complex.";  
 RL Science 255:306-312(1992).  
 RN [21]  
 RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS).  
 RX MEDLINE-95075462; PubMed:7984244;  
 RA Somers W., Ullsch M., de Vos A.M., Kossiakoff A.A.,  
 RT "The X-ray structure of a growth hormone-prolactin receptor complex.";  
 RL Nature 372:478-481(1994).  
 RN [22]  
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
 RA Chantalat S., Chirgadze N.Y., Jones N., Korber F., Navaza J.,  
 RA Pavlovsk A.G., Wlodawer A.,  
 RT "The crystal-structure of wild-type growth-hormone at 2.5-A  
 resolution.";  
 RL Protein Pept. Lett. 2:333-340(1995).  
 RN [23]  
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
 RX MEDLINE-97113023; PubMed:8943276;  
 RA Sundstroem M., Lundqvist T., Rosdin J., Giebel L.B., Millian D.,  
 RA Norstedt G.,  
 RT "Crystal structure of an antagonist mutant of human growth hormone,  
 GL20R, in complex with its receptor at 2.9-A resolution.";  
 RL J. Biol. Chem. 271:32197-32203(1996).  
 RN [24]  
 RP VARIANT CYS-165.  
 RX MEDLINE-99318093; PubMed:10391209;  
 RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,  
 RA Shaw N., Lane C.R., Lim E.P., Kalyanaram N., Nimesh J., Ziaugra L.,  
 RA Friedland L., Rolfe A., Warrington J., Lipshutz R., Baley G.O.,  
 RA Lander E.S.,  
 RT "Characterization of single-nucleotide polymorphisms in coding regions  
 of human genes.";  
 RL Nat. Genet. 22:231-238(1999).  
 RN [25]  
 RP ERRATUM.  
 RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,  
 RA Shaw N., Lane C.R., Lim E.P., Kalyanaram N., Nimesh J., Ziaugra L.,  
 RA Friedland L., Rolfe A., Warrington J., Lipshutz R., Baley G.O.,  
 RA Lander E.S.,  
 RL Nat. Genet. 23:373-373(1999).  
 CC -!- FUNCTION: Plays an important role in growth control. Its major  
 CC role in stimulating body growth is to stimulate the liver and  
 CC other tissues to secrete IGF-1. It stimulates both the  
 CC differentiation and proliferation of myoblasts. It also stimulates

```

Query Match          98.1%  Score 255;  DB 1;  Length 217;
Best Local Similarity 100.0%  Pred. No. 6.6e-25;
Matches 48;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY  2  FPTPLSLRFLDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONP 49
DB  27  FPTPLSLRFLDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONP 74

RESULT 2
SOMA_MACMO
ID  SOMA_MACMO  STANDARD:  PRT:  217 AA.
AC  P33093;
DT  01-OCT-1993 (Rel. 27, Created)
DT  01-OCT-1994 (Rel. 30, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  Somatotropin precursor (Growth hormone) (GH) (GH-N) (Pituitary growth
DE  hormone) (Growth hormone 1).
GN  GH1.
OS  Pan troglodytes (Hominidae).
OC  Eukaryota; Metazoa; Chordata; Primates; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.
OC  Cercopithecoidea; Macaca.
OX  NCBI_TaxID:9544;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE:94008724; PubMed:8404617;
RA  Golos T.G., Burning M., Fisher J.M., Fox R.P.D.;
RT  "Cloning of four growth hormone/chorionic somatomotropin-related
RT  complementary deoxyribonucleic acids differentially expressed during
RT  pregnancy in the rhesus monkey placenta."
RL  Endocrinology 133:1744-1752(1993).
RN  [2]
RP  SEQUENCE OF 27-217.
RX  MEDLINE:86129460; PubMed:3080959;
RA  Li C.H., Ching D., Lahm R.W., Stein S.;
RT  "The primary structure of monkey pituitary growth hormone."
RL  Arch. Biochem. Biophys. 245:287-291(1986).
CC  -!- FUNCTION: Plays an important role in growth control. Its major
CC  role in stimulating body growth is to stimulate the liver and
CC  other tissues to secrete IGF-1. It stimulates both the
CC  differentiation and proliferation of myoblasts. It also stimulates
CC  amino acid uptake and protein synthesis in muscle and other
CC  tissues.
CC  -!- SUBCELLULAR LOCATION: Secreted.
CC  -!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  EMBL: L16556; AAA18842.1;
CC  PIR: I67410; I67410.
CC  BSSP: p01241; IAXI.
DR  InterPro: IPR001400; Somatotropin.
DR  Pfam: PF00103; hormone; 1.
DR  PRINTS: PR00836; SOMATOTROPIN.
DR  PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR  PROSITE: PS00338; SOMATOTROPIN_2; 1.
KW  Hormone; Pituitary; Signal.
FT  SIGNAL 1 26
FT  CHAIN 27 217
FT  DISULFID 79 191
FT  DISULFID 208 215
FT  DISULFID 208 215
SQ  SEQUENCE 217 AA; 24913 MW; 2C5180341EEC46D0 CRC64;

Query Match          98.1%  Score 255;  DB 1;  Length 217;
Best Local Similarity 100.0%  Pred. No. 6.6e-25;
Matches 48;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY  2  FPTPLSLRFLDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONP 49
DB  27  FPTPLSLRFLDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONP 74

RESULT 4
SOMA_CALJA
ID  SOMA_CALJA  STANDARD:  PRT:  217 AA.
AC  Q9GMB3;
DT  28-FEB-2003 (Rel. 41, Created)
DT  28-FEB-2003 (Rel. 41, Last sequence update)

```

```

Best Local Similarity 100.0%  Pred. No. 6.6e-25;
Matches 48;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY  2  FPTPLSLRFLDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONP 49
DB  27  FPTPLSLRFLDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONP 74

RESULT 3
SOMA_PANTR
ID  SOMA_PANTR  STANDARD:  PRT:  217 AA.
AC  P58756;
DT  28-FEB-2003 (Rel. 41, Created)
DT  28-FEB-2003 (Rel. 41, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  Somatotropin precursor (Growth hormone) (GH) (GH-N) (Pituitary growth
DE  hormone) (Growth hormone 1).
GN  GH1.
OS  Pan troglodytes (Hominidae).
OC  Eukaryota; Metazoa; Chordata; Primates; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.
OC  Cercopithecoidea; Macaca.
OX  NCBI_TaxID:9598;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Revor A., Esquivel D., Santiago D., Barrera-Saldana H.;
RT  "Independent duplication of the growth hormone gene in three
RT  Anthropoid lineages."
RL  Submitted (APR-2001) to the EMBL/Genbank/DDBJ databases.
CC  -!- FUNCTION: Plays an important role in growth control. Its major
CC  role in stimulating body growth is to stimulate the liver and
CC  other tissues to secrete IGF-1. It stimulates both the
CC  differentiation and proliferation of myoblasts. It also stimulates
CC  amino acid uptake and protein synthesis in muscle and other
CC  tissues (by similarity).
CC  -!- SUBCELLULAR LOCATION: Secreted.
CC  -!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  EMBL: AF374232; AA172284.1;
CC  InterPro: IPR001400; Somatotropin.
CC  Pfam: PF00103; hormone; 1.
CC  PRINTS: PR00836; SOMATOTROPIN.
CC  PROSITE: PS00266; SOMATOTROPIN_1; 1.
CC  PROSITE: PS00338; SOMATOTROPIN_2; 1.
KW  Hormone; Pituitary; Signal.
FT  SIGNAL 1 26
FT  CHAIN 27 217
FT  DISULFID 79 191
FT  DISULFID 208 215
FT  DISULFID 208 215
SQ  SEQUENCE 217 AA; 24843 MW; FE245E5E0518674 CRC64;

Query Match          98.1%  Score 255;  DB 1;  Length 217;
Best Local Similarity 100.0%  Pred. No. 6.6e-25;
Matches 48;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY  2  FPTPLSLRFLDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONP 49
DB  27  FPTPLSLRFLDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONP 74

RESULT 4
SOMA_CALJA
ID  SOMA_CALJA  STANDARD:  PRT:  217 AA.
AC  Q9GMB3;
DT  28-FEB-2003 (Rel. 41, Created)
DT  28-FEB-2003 (Rel. 41, Last sequence update)

```



modified and this statement is not recommended. Usage by and for commercial entities requires a license agreement. See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

EMBL: AF374233; AAL72285.1; InterPro: IPR001400; Somatotropin. Pfam: PF00103; hormone; 1. PROSITE: PS00266; SOMATOTROPIN\_1; 1. PROSITE: PS00338; SOMATOTROPIN\_2; 1. Hormone; Placenta; Signal; Glycoprotein. FT SIGNAL 1 26 BY SIMILARITY. FT CHAIN 27 216 BY GROWTH HORMONE VARIANT. FT DISULFID 3 191 BY SIMILARITY. FT DISULFID 208 215 BY SIMILARITY. SQ SEQUENCE 217 AA; 24990 MW; 15924290756777DE CRC64;

Query Match 50.8%; Score 246; DB 1; Length 217; Best Local Similarity 93.8%; Pred No. 160; 22; Matches 45; Conservation 1; Indels 0; Gaps 0;

QY 2 PFTITSLPLFNAMIRAHRLHGLAFIVQ EAYTPKEQKYSFIQNP 49  
II  
DB 27 PFTITSLPLFNAMIRAHRLHGLAFIVQ EAYTPKEQKYSFIQNP 74

RESULT 7  
SOM2\_HUMAN STANDARD: PRT: 217 AA.  
AC P01242; P05987; 21-JUL-1986 (Rel. 01, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Growth hormone variant precursor (GH-V) (Placenta-specific growth hormone)  
DE hormone (Growth hormone 2).  
GN GH2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID:9606;  
RN [1]  
RX MEDLINE-8182010; PubMed-7169009;  
RA Seeburg P.H.;  
RT "The human growth hormone gene family: nucleotide sequences show recent divergence and predict a new polypeptide hormone.";  
RL DNA 1:239-249(1982).

[2]  
RN SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RX MEDLINE-88243769; PubMed-3379057;  
RA Cooke N.E., Ray J., Emery J.G., Lieber S.A.;  
RT "Two distinct species of human growth hormone variant mRNA in the human placenta predict the expression of novel growth hormone proteins.";  
RL J. Biol. Chem. 263:9001-9006(1988).  
RN [3]  
RX SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE-84624984; PubMed-2460050;  
RA Ignotz A., Scippo M.L., Frankenre F., Hennon G.;  
RT "Cloning and nucleotide sequence of placental hGH-v cDNA.";  
RL Arch. Int. Physiol. Biochim. 96:63-67(1988).  
RN [4]  
RX SEQUENCE FROM N.A.  
RX MEDLINE-89307277; PubMed-2744760;  
RA Chen E.Y., Liao Y.C., Smith D.H., Barrera-Saldana H.A.,  
RA Gelinas R.E., Seeburg P.H.;  
RT "The human growth hormone locus: nucleotide sequence, biology, and evolution.";  
RL Genomics 4:479-497(1989).  
RN [5]  
RX SEQUENCE FROM N.A.  
RC TISSUE: Placenta;  
RX MEDLINE-22388257; PubMed-12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Waqner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buelow K.H., Schaefer C.F., Hsieh N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.J., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Ronald M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Mazny D.M., Sodergren E., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";  
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [5]  
RN REVIEW.  
RX MEDLINE-99321812; PubMed-10393484;  
RA Baumann G.;  
RT "Growth hormone heterogeneity in human pituitary and plasma.";  
RL Horm. Res. 51 Suppl. 1:2-6(1999).  
CC -!- FUNCTION: Plays an important role in growth control. Its major role in stimulating body growth is to stimulate the liver and other tissues to secrete IGF-1. It stimulates both the differentiation and proliferation of myoblasts. It also stimulates amino acid uptake and protein synthesis in muscle and other tissues.  
CC -!- SUBUNIT: Monomer, dimer, trimer, tetramer and pentamer, disulfide-linked or non-covalently associated, in homopolymeric and heteropolymeric combinations. Can also form a complex either with GHBP or with the alpha<sub>2</sub>-macroglobulin complex.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event-Alternative splicing: Named isoforms=2;  
CC Name=1; Synonyms=GH-V1;  
CC IsoId=P01242-1; Sequence=Displayed;  
CC Name=2; Synonyms=GH-V2;  
CC IsoId=P01242-2; Sequence=VSP\_006203;  
CC Note=No experimental confirmation available;  
CC -!- TISSUE SPECIFICITY: Expressed in the placenta.  
CC -!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
EMBL: K0470; AAA98619.1;  
EMBL: J03756; AAB59547.1;  
EMBL: J03756; AAB59549.1;  
EMBL: M38451; AAA35891.1;  
EMBL: J03671; AAA52552.1;  
EMBL: BC020760; AAB20760.1;  
PIR: A28072; SIHUV2;  
PIR: D12435; SIHUV;  
HSSP: P01241; 1A22;  
DR Genew; HGNC:4262; GH2.  
DR MIN; 139240;  
DR GO:0005180; F:peptide hormone; TAS.  
DR InterPro: IPR001400; Somatotropin.  
DR Pfam: PF00103; hormone; 1.  
DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
KW Hormone; Placenta; Signal; Glycoprotein; Alternative splicing;  
KW Polymorphism.  
FT SIGNAL 1 26



DNA 1:239-249(1982).  
 RL SEQUENCE FROM N.A.  
 RN  
 RP  
 RA TISSUE=Placenta, and Uterus;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg K.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schale G.D.,  
 RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,  
 RA Slapleton M., Soares M.B., Bonaldo M.P., Casavani T.L., Schenck P.B.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Kana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mailahy S.J.,  
 RA Besac S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gillis R.A.,  
 RA Fahey C., Jellison E., Kettmann M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Yama A.C., Shengshenku Y., Bouffard J.B.,  
 RA Blackley A.W., Touchman J.W., Green E.D., Jackson R.,  
 RA Rodriguez A., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield J.S.N., Krzywicki M.L., Skasko M., Shalika D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RA \*Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.\*  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16904(2002).  
 RN [7].  
 RN SEQUENCE OF 50-217 FROM N.A.  
 RP MEDLINE=7607161; PubMed=593368;  
 RX Shine J., Seeburg P.H., Martial J.A., Baxter J.D., Goodman H.M.,  
 RA \*Construction and analysis of recombinant DNA for human chorionic  
 RT somatotropin.\*  
 RL Nature 270:494-499(1977).  
 RN [8].  
 RN SEQUENCE OF 27-217.  
 RP MEDLINE=73201971; PubMed=4712450;  
 RX Li C.H., Dixon J.S., Chung D.,  
 RA \*Amino acid sequence of human chorionic somatotropin.\*  
 RL Arch. Biochem. Biophys. 155:95-110(1973).  
 RN [9].  
 RN SEQUENCE OF 27-117.  
 RP MEDLINE=72016313; PubMed 5286363;  
 RX Sherwood L.M., Handwerger S., McLaurin W.D., Lerner M.,  
 RA \*Amino acid sequence of human placental lactogen.\*  
 RL Nature New Biol. 233:59-61(1971).  
 RN [10].  
 RP ERRATUM.  
 RA Sherwood L.M., Handwerger S., McLaurin W.D., Lerner M.,  
 RL Nature New Biol. 235:64-64(1972).  
 RN [11].  
 RP INTERCHAIN DISULFIDE BONDS.  
 RX Schneider A.B., Kowalski K., Russell J., Sherwood L.M.,  
 RA \*Identification of the interchain disulfide bonds of dimeric human  
 RT placental lactogen.\*  
 RL J. Biol. Chem. 254:3782-3787(1979).  
 CC -1- FUNCTION: SIMILAR TO THAT OF SOMATOTROPIN.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- MISCELLANEOUS: THE SEQUENCE OF CSH1 IS SHOWN.  
 CC -1- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL: V00573; CAA23836.1;  
 DR EMBL: J00289; AAA98747.1;  
 DR EMBL: K02401; AAA52115.1;  
 DR EMBL: M15894; AAA52116.1;  
 DR EMBL: J03071; AAA52551.1;

DR EMBL: J06118; AAA98621.1;  
 DR EMBL: BC002717; AAH02717.1;  
 DR EMBL: BC005921; AAH05921.1;  
 DR EMBL: BC020756; AAH20756.1;  
 DR PIR: A26449; A26449.  
 DR PIR: C32435; LCHUC.  
 DR HSSP: P01241; 1A22.  
 DR Gene: HGNC:2440; CSH1.  
 DR Gene: HGNC:2441; CSH2.  
 DR MIM: 150200;  
 DR GO: GO:0007565; P:pregnancy; TAS.  
 DR InterPro: IPR01400; Somatotropin.  
 DR Pfam: PF00103; Hormone; 1.  
 DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
 DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
 KW Hormone; Placenta; Multidomain family; Signal.  
 FT SIGNAL: 1 26  
 FT CHAIN: 27 217  
 FT DISULFID: 75 141  
 FT DISULFID: 208 215  
 FT DISULFID: 208 208  
 FT DISULFID: 215 215  
 FT VARIANT: 3 3  
 FT VARIANT: 104 105  
 FT CONFLICT: 84 84  
 FT CONFLICT: 95 95  
 FT CONFLICT: 116 116  
 FT CONFLICT: 134 136  
 FT CONFLICT: 134 136  
 FT SEQUENCE: 217 AA; 25020 MW; 23580D27A713F431 CRC64;  
 Query Match 75.8%; Score 197; DB 1; Length 217;  
 Rest Local Similarity 80.0%; Pred. No. 130-17;  
 Matches 36; Conservative 5; Mismatches 4; Indels 0; Gaps 0;  
 CY 4 TIPLSKLEFDNMLRAHPLHQLAFYQFFEFAYIPKQKYSFLHD 48  
 DB 29 TVPLSRLEFDHMLQAHRAHQLALDTYQFFETYPKQKYSFLHD 73  
 RESULT 10  
 SOMA\_MOUSE  
 ID SOMA\_MOUSE STANDARD: PRT: 215 AA.  
 AC P06880;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, last sequence update)  
 DT 28-FEB-2003 (Rel. 41, last annotation update)  
 DE Somatotropin precursor (Growth hormone).  
 GN GH1 OR GH  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chorda a; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85261358; PubMed=2091252;  
 RA Linzer D.L.H., Talamantes F.;  
 RT \*Nucleotide sequence of mouse prolactin and growth hormone mRNAs and  
 RL expression of these mRNAs during pregnancy.\*  
 RL J. Biol. Chem. 260:9574-9579(1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=F27DU; TISSUE: Liver;  
 RX MEDLINE=96194803; PubMed=847448;  
 RA Das P., Meyer L., Seyfert L.-M., Brockmann G., Schwerin M.;  
 RT \*Structure of the growth hormone encoding gene and its promoter in  
 RT mice.\*  
 RL Gene 169:209-213(1996).  
 CC -1- FUNCTION: Plays an important role in growth control. Its major  
 CC role in stimulating body growth is to stimulate the liver and  
 CC other tissues to secrete IGF-1. It stimulates both the  
 CC differentiation and proliferation of myoblasts. It also stimulates

CC amino acid uptake and protein synthesis in muscle and other  
 CC tissues.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed, usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: X02891; CAA26650.1;  
 CC EMBL: Z46663; CAA86658.1;  
 CC PIR: B23911; STMS.  
 CC HSSP: P01246; 1BST.  
 CC MDL: M31395-267; GRI.  
 CC InterPro: IPR001400; Somatotropin.  
 CC Pfam: PF00103; hormone; 1.  
 CC PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
 CC Hormone; Pituitary; Signal.  
 CC SIGNAL 1 26 BY SIMILARITY.  
 CC CHAIN 27 216 SOMATOTROPIN.  
 CC DISULFID 78 189 BY SIMILARITY.  
 CC DISULFID 206 214 BY SIMILARITY.  
 CC SEQUENCE 216 AA: 24716 MW: 9866kDaAE25D65FC CRC64:  
 Query Match 62.18; Score 161.5; DB 1; Length 21;  
 Best Local Similarity 68.18; Pred. No. 3.6e-13;  
 Matches 42; Conservative 6; Mismatches 8; Indels 1; Gaps 1;  
 QY 2 FPTPLSRFDNMLRAHRLHQIAFYTYQFFRAYIPKEQKYSFLON 48  
 DB 27 FPAMPLSSLESNVLAHRLHQIAFYTYQFFRAYIPKEQKYS-ION 72  
 RESULT 1;  
 ID SOMA\_BALBO STANDARD: PRT: 190 AA.  
 AC P31092;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Somatotropin (Growth hormone).  
 GN GRI.  
 OS Balanoptera borealis (Sei whale).  
 OC Eukaryota; Metazoa; Chordata; Cetartiodactyla; Cetacea; Mysticeti;  
 OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;  
 OC Balanopteridae; Balanoptera  
 CC NCBI TaxID: 4768;  
 RN 1;  
 RP SEQUENCE;  
 RX MFLINSE-B00-0569; PubMed-2115614;  
 RA Yudaev N.A., Pankov Y.A., Bulatov A.A., Osipova T.A.;  
 RT "Amino acid sequence of seal whale somatotropin."  
 RL Biochimica 47:1059-1069(1982).  
 RN [2]  
 RP PRELIMINARY PARTIAL SEQUENCE;  
 RA Osipova T.A., Bulatov A.A., Pankov Y.A.;  
 RT "Structural studies of tryptic peptides from large cyanogen bromide  
 RI fragments of sei whale (Balanoptera borealis) somatotropin."  
 RL Bioorg. Khim. 4:1589-1599(1978).  
 CC -!- FUNCTION: Plays an important role in growth control. Its major  
 CC role in stimulating body growth is to stimulate the liver and  
 CC other tissues to secrete IGF-1. It stimulates both the  
 CC differentiation and proliferation of myoblasts. It also stimulates  
 CC amino acid uptake and protein synthesis in muscle and other  
 CC tissues.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.  
 CC PIR: PN0140; PN0140.

DR HSSP: P01241; 1AXI.  
 DR InterPro: IPR001400; Somatotropin.  
 DR Pfam: PF00103; hormone; 1.  
 DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
 DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
 DR KW Hormone; Pituitary.  
 DR FT DISULFID 52 163 BY SIMILARITY.  
 DR FT DISULFID 180 188 BY SIMILARITY.  
 DR SQ SEQUENCE 190 AA: 21835 MW: 9936kDaB14A75D6 CRC64:  
 Query Match 61.78; Score 160.5; DB 1; Length 190;  
 Best Local Similarity 68.18; Pred. No. 4.1e-13;  
 Matches 32; Conservative 6; Mismatches 8; Indels 1; Gaps 1;  
 QY 2 FPTPLSRFDNMLRAHRLHQIAFYTYQFFRAYIPKEQKYSFLON 48  
 DB 1 FPAMPLSSLESNVLAHRLHQIAFYTYQFFRAYIPKEQKYS-ION 46  
 RESULT 1;  
 ID SOMA\_VULVU STANDARD: PRT: 190 AA.  
 AC P10766;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Somatotropin (Growth hormone).  
 GN GRI.  
 OS Loxodonta africana (African elephant).  
 OC Eukaryota; Metazoa; Chordata; Cetartiodactyla; Cetacea; Proboscidea; Loxodontidae; Loxodonta.  
 CC NCBI TaxID: 9785;  
 RN 1;  
 RP SEQUENCE;  
 RA Holmes J.D., Miedel M.C., Li C.H., Pan Y.C.E.;  
 RT "Primary structure of elephant growth hormone."  
 RL Int. J. Pept. Protein Res. 33:368-372(1989).  
 CC -!- FUNCTION: Plays an important role in growth control. Its major  
 CC role in stimulating body growth is to stimulate the liver and  
 CC other tissues to secrete IGF-1. It stimulates both the  
 CC differentiation and proliferation of myoblasts. It also stimulates  
 CC amino acid uptake and protein synthesis in muscle and other  
 CC tissues.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.  
 CC PIR: JK0219; JK0219.  
 DR HSSP: P01246; 1BST.  
 DR InterPro: IPR001400; Somatotropin.  
 DR Pfam: PF00103; hormone; 1.  
 DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
 DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
 DR KW Hormone; Pituitary.  
 DR FT DISULFID 52 163 BY SIMILARITY.  
 DR FT DISULFID 180 188 BY SIMILARITY.  
 DR SQ SEQUENCE 190 AA: 21761 MW: 0586kDaB081A0B741F2 CRC64:  
 Query Match 61.38; Score 159.5; DB 1; Length 190;  
 Best Local Similarity 68.18; Pred. No. 5.5e-13;  
 Matches 32; Conservative 6; Mismatches 8; Indels 1; Gaps 1;  
 QY 2 FPTPLSRFDNMLRAHRLHQIAFYTYQFFRAYIPKEQKYSFLON 48  
 DB 1 FPAMPLSSLESNVLAHRLHQIAFYTYQFFRAYIPKEQKYS-ION 46  
 RESULT 1;  
 ID SOMA\_VULVU STANDARD: PRT: 190 AA.  
 AC P10766;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Somatotropin (Growth hormone).



```

CC -1- FUNCTION: Plays an important role in growth control. Its major
CC role in stimulating body growth is to stimulate the liver and
CC other tissues to secrete IGF-1. It stimulates both the
CC differentiation and proliferation of myoblasts. It also stimulates
CC amino acid uptake and protein synthesis in muscle and other
CC tissues.
CC
CC -1- SUBCELLULAR LOCATION: Secreted.
CC
CC -1- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: 025973; AAA67294.1;
CC EMBL: 013390; AAA36142.1;
CC PIR: J08642; J04642;
CC HSP: P01246; IRS1.
CC
CC InterPro: I06031400; Somatotropin.
CC
CC Pfam: PF06103; hormone.1;
CC
CC PROSITE: PS00266; SOMATOTROPIN_1; 1;
CC PROSITE: PS00338; SOMATOTROPIN_2; 1;
CC
CC KW Hormone; Pituitary; Signal;
CC
CC FT SIGNAL 1 26 BY SIMILY Y.
CC
CC FT CHAIN 27 216 SOMATOTR I.
CC
CC FT DISULFID 78 189 BY SIMILY Y.
CC
CC FT DISULFID 206 214 BY SIMILIAITY.
CC
CC FT CONFLICT 7 7 N -> T (1 REF. 2).
CC
CC FT CONFLICT 26 26 T -> A (1 REF. 2).
CC
CC FT CONFLICT 159 159 G -> A (1 REF. 2).
CC
CC FT CONFLICT 181 181 L -> P (1 REF. 2).
CC
CC SQ SEQUENCE 216 AA; 24454 MW; 05823239A7D292C6 CRC64;
CC
CC Query Match: 61.3%; Score 159.5; DB 1; Length 216;
CC Best Local Similarity 68.1%; Pred.No 6,3e-13;
CC Matches 32; Conservative 6; Mismatches 8; Indels 1; Gaps 1;
CC
CC QY 2 FPTIPISRLFDNAMLRAHRIHQIAFDITYGFEEAYIPKEQKYSFLQN 48
CC
CC Db 27 FPMPLSSFLANVLRAQHIHQIAADTYKFEFRAYIPGEQKYSIQN 72

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:54:30 : Search time 19.1434 Seconds  
(without alignments)  
660,520 Million cell updates/sec

Title: US-09-423-100-1

Perfect score: 260

Sequence: 1 MFPTPLSLFDNMLRAHR.....QPEEAYIPKESQKSYFLQNP 49

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 3.5

Searched: 840525 seqs, 25805264 residues

Total number of hits satisfying chosen parameters: 840525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_23:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_ordanelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_virus:\*

16: sp\_bacteriophage:\*

17: sp\_archaeal:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result NO	Score	% Match	Length	DB ID	Description
1	249	95.8	217	6 Q8WNE0	Q8WNE0 ateles geof
2	228	87.7	245	4 O14644	O14644 homo sapien
3	214	81.9	212	6 Q07468	Q07468 macaca mula
4	213	81.9	217	6 Q07467	Q07467 macaca mula
5	201	77.3	217	6 Q07469	Q07469 macaca mula
6	197	75.8	217	4 O14407	O14407 homo sapien
7	195	75.0	217	6 Q8WNE9	Q8WNE9 ateles geof
8	170	65.4	217	6 O8M174	O8M174 callithrix
9	160.5	61.7	216	11 Q70615	Q70615 spalax leuc
10	159.5	61.3	52	6 Q9TV91	Q9TV91 equus cabal
11	159.5	61.3	216	6 O8M173	O8M173 delphinus d
12	159.5	61.3	216	6 Q8HYE5	Q8HYE5 alluropoda
13	155.5	59.8	216	11 Q9R2C3	Q9R2C3 mus musculus
14	154	59.2	216	11 Q9JMK4	Q9JMK4 cavia porce
15	152	58.5	178	6 Q95MJ5	Q95MJ5 tarsius ban
16	149.5	57.5	204	6 Q95205	Q95205 ovus aries

17	147	56.5	202	4 O14643	O14643 homo sapien
18	146	56.2	178	6 Q95MJ6	Q95MJ6 tarsius syr
19	144	55.4	190	11 Q9JJK0	Q9JJK0 cavia porce
20	143.5	55.2	192	6 Q9TJ21	Q9TJ21 capra hircu
21	142.5	54.9	192	6 Q9TOW9	Q9TOW9 bos indicus
22	141	54.2	217	6 Q8M175	Q8M175 callithrix
23	138.5	53.3	217	6 Q28957	Q28957 sus scrofa
24	138.5	53.3	217	6 Q9BEC0	Q9BEC0 traquilus ja
25	138.5	53.3	217	6 Q9BEB9	Q9BEB9 traquilus ja
26	137	52.7	40	6 Q9TR19	Q9TR19 macropus ru
27	132	50.8	218	13 Q9P072	Q9P072 cynops pyrri
28	128	49.2	195	13 Q91386	Q91386 amia calva
29	123.5	47.5	143	6 Q95240	Q95240 canis fami
30	122	46.9	199	4 Q14406	Q14406 homo sapien
31	119	45.8	63	13 Q8QGB5	Q8QGB5 anser anser
32	107	41.2	53	6 Q19034	Q19034 ovus aries
33	104	40.0	55	6 Q46474	Q46474 telis silve
34	101	38.8	167	4 Q98451	Q98451 homo sapien
35	101	38.8	200	13 Q8QFM8	Q8QFM8 clarias bat
36	99.5	38.3	145	6 Q9B8R4	Q9B8R4 galago cras
37	99	38.3	200	13 Q8QGB1	Q8QGB1 heteropneus
38	99	38.3	200	13 Q8QFN2	Q8QFN2 clarias gar
39	96	36.9	50	13 Q9P8K5	Q9P8K5 clarias gar
40	95	36.5	200	13 Q8AV82	Q8AV82 silurus aso
41	90	34.6	188	13 Q90283	Q90283 carassius a
42	87	33.5	188	13 Q8UVP2	Q8UVP2 cirrhinus m
43	87	33.5	188	13 Q90W26	Q90W26 carassius a
44	87	33.5	188	13 Q90W27	Q90W27 carassius a
45	87	33.5	188	13 Q98SK7	Q98SK7 cyprinus ca

## ALIGNMENTS

RESULT 1

Q8WNE0 Q8WNE0 PRELIMINARY: PRT: 217 AA.  
AC Q8WNE0:  
DI 01-MAR-2002 (TRENBLrel. 20, Created)  
DI 01-MAR-2002 (TRENBLrel. 20, Last sequence update)  
DI 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
DE Growth hormone.  
GN GH-N.  
OS Ateles geoffroyi (Black-handed spider monkey).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Ateleinae; Ateles.  
RX NCBI\_TaxID=9509;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA Revol A., Esquivel D., Santiago D., Barrera-Saldana H.;  
RT \*Independent duplication of the growth hormone gene in three  
RT Anthropoledean lineages\*;  
RL Submitted (APR-2001) to the EMBL/GenBank/DDIJ databases.  
DR EMBL: AF374234; AAL72286.1;  
DR InterPro: IPR001400; Somatotropin.  
DR Pfam: PF00103; hormone; 1.  
DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
SQ SEQUENCE 217 AA: 24894 MW: 425829FF41EEAAE6 CRC64:

Query Match 95.8% Score 249; ID 6; Length 217;

Best Local Similarity 97.9%; Pred. No. 1.2e-25;

Matches 47; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 FPTPLSLFDNMLRAHR.HOLAFDTYQEFEEAYIPKESQKSYFLQNP 49

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 27 FPTPLSLRLDNLMLRA.HLHOLAFDTYQEFEEAYIPKESQKSYFLQNP 74

RESULT 2

O14644 PRELIMINARY: PRT: 245 AA.  
ID O14644  
AC O14644;

Query Match 77.3%; Score 201; DB 6; Length 217;  
Best Local Similarity 74.5%; Pred. No. 3.8e-19;

```

Matches 35; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

QY 3 PTPLSRFLDNAMLRHRLHQLAFDTYQFEFEAVIPKEQKYSFLQNP 49
   1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:|||||
DB 28 PSVPLSRFLDNIMQAHRLHQLAFDTYQFEFEKTYIPKEKKISLQNP 74

RESULT 6
ID Q14407 PRELIMINARY: PRT: 217 AA.
AC Q14407
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Chorionic somatomotropin CS-2 (Chorionic somatomotropin hormone
DE 2).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Ateles.
OX NCBI_TaxID=9606;
RN SEQUENCE FROM N.A.
RP MEDLINE=89307277; PubMed=2744760;
RA Chen E.Y., Liao Y.C., Smith D.H., Barrera-Saldana H.A., Gelinas R.F.,
RA Seeburg P.H.
RT "The human growth hormone locus: nucleotide sequence, biology, and
RT evolution".
RL Science 252:1745-1748(1990).
RN 13;
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
RI Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN 14;
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: J03071; AA525531;
DR EMBL: H022044; AA822044;
DR EMBL: H035965; AA835965;
DR HSSP: P01241; 1A22.
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone; 1.
DR PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00348; SOMATOTROPIN_2; 1.
SQ SEQUENCE 217 AA; 2494 MW; 39FAA10B66B2E551 CRC64.

Query Match 75.8%; Score 147; DB 4; Length 217;
Best Local Similarity 80.0%; Pred. No. 1; 18;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 4 TPLSRFLDNAMLRHRLHQLAFDTYQFEFEAVIPKEQKYSFLQNP 48
   1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:|||||
DB 24 TVPLSRFLDNAMLRHRLHQLAFDTYQFEFEKTYIPKEKKISLQNP 73

RESULT 7
ID Q6ND9 PRELIMINARY: PRT: 217 AA.
AC Q6ND9;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Growth hormone.

```

```

GN GH-V.
OS Ateles geoffroyi (Black-handed spider monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Ateles.
OX NCBI_TaxID=9509;
RN SEQUENCE FROM N.A.
RA Revol A., Esquivel D., Santiago D., Barrera-Saldana H.;
RT "Independent duplication of the growth hormone gene in three
RT Anthropoidean lineages".
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF374235; AAL72287.1;
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone; 1.
DR PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00348; SOMATOTROPIN_2; 1.
UR PROSITE: PS00348; SOMATOTROPIN_2; 1.
SQ SEQUENCE 217 AA; 2523 MW; 741745A13767054E FNC64;

Query Match 75.0%; Score 195; DB 6; Length 217;
Best Local Similarity 77.1%; Pred. No. 2; 50; 14;
Matches 37; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 2 FPTPLSRITINAMLRHRLHQLAFDTYQFEFEAVIPKEQKYSFLQNP 49
   1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:|||||
DB 27 FPRIPLSKRLGDAMLRHRLHQLAFDTYQFEFEAVIPKEQKYSFLQNP 74

RESULT 8
ID Q8M174 PRELIMINARY: PRT: 217 AA.
AC Q8M174;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Growth hormone-like protein 6 precursor.
GN GHP6.
OS Callithrix jacchus (Common marmoset).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Callitrich.
OX NCBI_TaxID=9483;
RN SEQUENCE FROM N.A.
RA Wallis D.C., Wallis M.;
RT "Characterisation of the GII gene cluster in a new-world monkey, the
RT marmoset (Callithrix jacchus)".
RL J. Mol. Endocrinol. 0:0-0(2002).
DR EMBL: AJ489811; CAJ34012.1;
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone; 1.
DR PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00348; SOMATOTROPIN_2; 1.
KW SIGNAL.
FT CHAIN 1 26 POTENTIAL.
FT CHAIN 27 217 GROWTH HORMONE LIKE PROTEIN 6.
SQ SEQUENCE 217 AA; 25177 MW; 5ECF148798278F1A CRC64;

Query Match 65.4%; Score 170; DB 6; Length 217;
Best Local Similarity 68.1%; Pred. No. 6; 10; 15;
Matches 32; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 3 PTPLSRFLDNAMLRHRLHQLAFDTYQFEFEAVIPKEQKYSFLQNP 49
   1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:||||| 1:|||||
DB 28 PTPLSKRLGDAMLRHRLHQLAFDTYQFEFEKTYIPKEKKISLQNP 74

RESULT 9
ID Q70615 PRELIMINARY: PRT: 216 AA.
AC Q70615;
DT 01-AUG-1998 (TRENBLrel. 07, Created)
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)

```

DE	Growth hormone precursor.
OS	Spalax leucodon ehrenbergi (Ehrenberg's mole rat).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Spalacinae;
OC	Nannospalax.
OX	NCBI_TaxId=30637;
RN	[[]]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=99124645; PubMed=9924177;
RA	Litoupis A., Nevo E., Wallis M.
RT	"Cloning and characterisation of the gene encoding mole rat (Spalax
RT	ehrenbergi) growth hormone."
RI	J. Mol. Endocrinol. 22:29-36(1999).
KL	EMBL: AJ005819; CAA06716.1;
DR	HSSP: P01241; IAXI
DR	InterPro: IPR01460; Somatotropin.
DR	Prm; Pituitary hormone; 1.
DR	PRINTS: PPR0347; SOMATOTROPIN.
DR	PROSITE: PS00428; SOMATOTROPIN 1.
DR	PROSITE: PS00448; SOMATOTROPIN_2; 1
KW	Signal.
FT	CHAIN 1 26
FT	CHAIN 27 216
FT	POTENTIAL..
FT	GROWTH HORMONE.
SQ	SEQUENCE 216 AA; 24627 HW; EEABRA133HAAADFE CRC64;
Query Match	61.7% Score 160.5 DB 1: Length 216;

```

Query Match      61.7%  Score 160.5  DB 1:  Length 216
Best local Similarity 68.1%  pred. No. 1,24 13
Matches 42  conservative 6  Mismatch 8  Indels 12

```

ID	PRELIMINARY:	PRT:	52 AA:
Q9TVV4			
AC	Q9TVV4		
DI	01 MAY-2000 (TrEMBLrel. 13, Created)		
DT	01 MAY 2000 (TrEMBLrel. 13, Last sequence update)		
DT	01-MAR 2003 (TrEMBLrel. 23, last annotation update)		
DE	Growth hormone (Fragment)		
DE	Growth hormone (Fragment)		
GN	GN		
OS	Equus caballus (Horse)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Perissodactyla; Equidae; Equus		
OX	NCBI_TaxID=9796;		
RI	[1]		
RP	SEQUENCE FROM N.A.		
RR	MEDLINE: 39160468; PubMed 10551423;		
RX	Castano A.R., Pomp D., Murray J.D., Howling A.T.		
RT	*Comparative mapping of 18 equine type I genes assigned by somatic		
RT	cell hybrid analysis.		
RL	Mamm. Genome 10:271-276(1994).		
DR	EMBL: AF037589; AAD2592.1; ..		
DR	HSSP: P01241; 1HGU.		
DR	InterPro: IPR001400; Somatotropin.		
DR	Pfam: PF00103; hormone, 1.		
FT	NON_TER 1		
FT	NON_TER 52		
FT	NON_TER 52		
SQ	SEQUENCE 52 AA: 20AGPVE[1]0FG6FA CRC64;		

	Query Match	61.3%	Score 154.5	DB 6	Length 52
	Best local Similarity	68.1%	Prod. No. 5.2e-14		
	Matches	42	Conservative	8	Indels 1
					Gaps
QY	2	FTPLSLFONAMLRHRIHQLAFDTYQEFEEAYIPKQKYSFLQN	48		
DB	6	FPMPLSLFLFANVLRAGHLLHQAADTYKEFERAYIPGEGYYS-ION	51		

RESULT 11  
Q8MI73

QBM173	PRELIMINARY;	PRT;	216 AA.
AC	QBM173;		
DT	01-OCT-2002 (T:EMBLrel. 22, Created)		
DT	01-OCT-2002 (T:EMBLrel. 22, Last sequence update)		
DT	01-MAR-2003 (T:EMBLrel. 23, Last annotation update)		
DE	Growth hormone precursor.		
DE	GH.		
OS	Delphinus delphis (Saddleback dolphin) (Black sea dolphin).		
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti; Delphinidae;		
OC	Delphinus.		
OX	NCBI_TaxID=9728;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE Layer;		
RA	Manion Z., Wallis O.C., Wallis M.;		
RT	*Cloning and characterisation of the cDNA from the common dolphin		
RT	(*Delphinus delphis) *;		
RL	Submitted (JUN-92) to the EMBL/Genbank/and Databases.		
UK	EMBL; A492191; C0037292.1;		
UK	Interfero: PR001400; Somatotropin.		
DR	Pfam: PF00103; Hormone; 1.		
DR	PROSITE: PR00836; SOMATOTROPIN.		
DR	PROSITE: PS00266; SOMATOTROPIN_1; 1.		
DR	PROSITE: PS00336; SOMATOTROPIN_2; 1.		
KW	SIGNAL.		
FT	SIGNAL. 1 26 POTENTIAL.		
FT	CHAIN 27 216 GROWTH HORMONE.		
SQ	SEQUENCE 216 AA; 24509 MW; 1FC457A84CFFEH02 CR:64;		

Query Match	61.3%	Score 159.5	RA 6	Length 216
Best Local Similarity	64.1%	Pred. No. 1.66	1%	
Matches 42	Conservative 6	Mismatches 8	Indels 1	Gaps 1

```

RESULT 12
QBHYE5
ID QBHYE5 PRELIMINARY: PRT: 216 AA.
AC QBHYE5
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Growth hormone precursor
OS Alluropoda melanoleuca (Giant panda).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Ursidae; Alluropoda.
OC NCBI Taxid:9646;
RN [1]
RS SEQUENCE FROM N.A.
RC TISSUE:Pituitary;
RC Liao M., Zhu M., Zhang A.;
RT "Cloning and expression of cDNA encoding growth hormone from
RT Alluropoda melanoleuca.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF540936; AAN77228.1;
KW Signal;
FT SIGNAL: 1 26 POTENTIAL.
FT SEQUENCE: 216 AA; 24483 MW; 444E017E44BCB56 CAC64;
SC

```

```

Query Match: 61.4% Score 159.5; B3 6; Length 216;
Best Local Similarity 58.7% Prod. No. 1.6e 14;
Matches 42; Conservative 6; Mismatches 8; Indels 1; Gaps 1;

Qy 2 FPTIPSLRFLDNLRAHRLHQLAFUTYOFFEAYIPKEOKYSLQN 48
    1 111 111 111 111 111 111 111 111 111 111 111
db 27 FPAFLSSLLCANVLRARHVLHGLAADTYKEFEKRAYIPGEKOKY-ION 72
    1 111 111 111 111 111 111 111 111 111 111 111

```

RESULT 13

```

Q9R2C3
ID Q9R2C3 PRELIMINARY: PRT: 216 AA.
AC Q9R2C3
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Growth hormone.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Nguyen T.N.K., Lieberhaber S.A.
RT "Mouse Growth Hormone Locus: Nucleotide Sequence and Phylogenetic
RT Analysis."
RL EMBL: U34462; AAC99988.1;
DR HSSP: P01241; IAXI.
DR InterPro: IPR001400; Somatotropin
DR Pfam: PF00103; hormone; 1.
DR PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
SQ SEQUENCE 216 AA: 24682 MW: 45968B19080D3 CRC64:

Query Match 59.8%; Score 15; DB 11; Length 216;
Best Local Similarity 66.0%; Pred. No. 6e-13;
Matches 31; Conservative 6; Mismatches 9; Indels 1; Gaps 1;

QY 2 FPTPLSLRFDNAMLRAHRLHQLAFDTYQEEAYIPKEQKYSFLOH 48
Db 27 FPMPLSLRFDNAMLRAHRLHQLAADTYKLEAYIPKEQKYSFLOH 72

RESULT 14
Q9JKM4
ID Q9JKM4 PRELIMINARY: PRT: 216 AA.
AC Q9JKM4
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Growth hormone precursor.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RA Odorico D.M., Fuller P.J., Herington A.C.
RT "Cloning and sequence of guinea pig growth hormone (GH).";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBCLS databases.
DR EMBL: AF234853; AAF36409.1;
DR HSSP: P01241; IAXI.
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone; 1.
DR PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
DR Signal.
KW SIGNAL.
FT CHAIN 27 216 POTENTIAL
FT SIGNAL.
SQ SEQUENCE 216 AA: 24822 MW: 45996B19080D3 CRC64:

Query Match 59.2%; Score 154; DB 11; Length 216;
Best Local Similarity 67.4%; Pred. No. 8.9e-13;
Matches 29; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

```

```

RESULT 15
Q95MJ5
ID Q95MJ5 PRELIMINARY: PRT: 178 AA.
AC Q95MJ5
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Growth hormone (Fragment).
OS Tarsius bancanus (Western tarsier) (Horsfield's tarsier).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Tarsiidae; Tarsiidae; Tarsiidae.
OX NCBI_TaxID=9477;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-21265430; PubMed-11371582;
RA Liu J.C., Makova K.D., Adkins R.M., Gibson S., Li W.H.
RT "Episodic Evolution of Growth Hormone in Primates and Emergence of the
RT Species Specificity of Human Growth Hormone Receptor."
RL Mol. Biol. Evol. 18:945-953 (2001).
DR EMBL: AF390611; AAK62408.1;
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone; 1.
DR PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
DR NON_TER 178
FT CHAIN 178
SQ SEQUENCE 178 AA: 20038 MW: F678087512F9H7F0 CRC64:

Query Match 58.5%; Score 152; DB 6; Length 178;
Best Local Similarity 65.1%; Pred. No. 1.4e-12;
Matches 28; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 2 FPTPLSLRFDNAMLRAHRLHQLAFDTYQEEAYIPKEQKYS 44
Db 6 FPSVPLSKLFANAVLRAOHLHQLAADTYKLEAYIPKEQKYS 50

Search completed: September 15, 2003, 12:03:27
Job time : 22.1434 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:44:15 : Search time 64.6308 Seconds  
(without alignments)  
225,942 Million cell updates/sec

Title: US-09-423-100-2  
Perfect score: 470  
Sequence: 1 MEPTIPLSRFDNAHLRAHR.....NLELRISLLILISWLEPVQ 52

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1167863 seqs, 154274573 residues

Total number of hits satisfying chosen parameters: 1000000

Minimum hit seq length: 5

Maximum hit seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*

1:	/SIDS1/qcqdta/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDS1/qcqdta/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDS1/qcqdta/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDS1/qcqdta/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDS1/qcqdta/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDS1/qcqdta/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDS1/qcqdta/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDS1/qcqdta/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDS1/qcqdta/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDS1/qcqdta/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDS1/qcqdta/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDS1/qcqdta/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDS1/qcqdta/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDS1/qcqdta/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDS1/qcqdta/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDS1/qcqdta/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDS1/qcqdta/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDS1/qcqdta/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDS1/qcqdta/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDS1/qcqdta/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDS1/qcqdta/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDS1/qcqdta/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SIDS1/qcqdta/geneseq/geneseq-emb1/AA2002.DAT.*
24:	/SIDS1/qcqdta/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY

Result NO.	Score	Query Match	Length DB	ID	Description
1	470	100.0	92	20	AA19856 Human growth hormo
2	470	100.0	134	20	AA192265 Human anti-angioten
3	470	100.0	150	20	AA192861 Chimeric protein,
4	465	98.9	140	10	AA191041 Human growth hormo
5	465	98.9	192	10	AA190129 Human growth hormo
6	465	98.9	192	20	AA192264 Human anti-angioten
7	465	98.9	261	10	AA191299 Human nerve growth
8	465	98.9	262	12	AA191740 Human growth hormo
9	465	98.9	310	11	AA193255 Fusion protein of

10	462	98.3	144	11	AA195313 Segment of B-cell
11	462	98.3	262	7	AA1961033 Human beta-nerve g
12	460	97.9	138	9	AA1981226 Sequence of protei
13	460	97.9	191	18	AA1920110 Primary amino acid
14	460	97.9	191	20	AA195809 Natural human 22kD
15	460	97.9	191	20	AA1904396 Mutant human 22kDa
16	460	97.9	191	20	AA1904397 Human growth hormo
17	460	97.9	191	21	AA1978425 Human growth hormo
18	460	97.9	191	22	AA197485 Human growth hormo
19	460	97.9	191	22	AA197486 Human growth hormo
20	460	97.9	191	23	AA1984859 Human growth hormo
21	460	97.9	191	23	AA1984860 Human growth hormo
22	460	97.9	191	23	AA1984861 Human growth hormo
23	460	97.9	191	23	AA1984862 Human growth hormo
24	460	97.9	191	23	AA1984863 Human growth hormo
25	460	97.9	191	23	AA1984864 Human growth hormo
26	460	97.9	191	23	AA1984907 Human growth hormo
27	460	97.9	191	23	AA1984908 Human growth hormo
28	460	97.9	191	23	AA1984909 Human growth hormo
29	460	97.9	191	23	AA1984910 Human growth hormo
30	460	97.9	191	23	AA1984911 Human growth hormo
31	460	97.9	191	23	AA1984912 Human growth hormo
32	460	97.9	191	23	AA1984913 Human growth hormo
33	460	97.9	191	23	AA1984914 Human growth hormo
34	460	97.9	191	23	AA1984915 Human growth hormo
35	460	97.9	191	23	AA1984916 Human growth hormo
36	460	97.9	191	23	AA1984917 Human growth hormo
37	460	97.9	191	23	AA1984918 Human growth hormo
38	460	97.9	191	23	AA1984919 Human growth hormo
39	460	97.9	191	23	AA1984920 Human growth hormo
40	460	97.9	191	23	AA1984921 Human growth hormo
41	460	97.9	191	23	AA1984922 Human growth hormo
42	460	97.9	191	23	AA1984923 Human growth hormo
43	460	97.9	191	23	AA1984924 Human growth hormo
44	460	97.9	191	23	AA1984925 Human growth hormo
45	460	97.9	191	23	AA1984926 Human growth hormo

ALIGNMENTS

RESULT 1  
AA192856  
ID AA192856 standard: protein: 92 AA.  
XX  
AC AA192856:  
XX  
DT 19-JAN-2000 (first entry)  
XX  
DE Human growth hormone (HGH) N terminal fragment #2.  
XX  
KW Growth hormone; chaperone; intramolecular; insulin; precursor;  
KW folding; conformation; chimeric protein; cleavable; recombinant;  
KW production; yield.  
XX  
CS Homo sapiens.  
XX  
PN WO/95/0102 A1.  
XX  
PD 07-OCT-1999.  
XX  
PF 31-MAR 1998; 98WO/000052.  
XX  
PR 31-MAR-1998; 98WO/000052.  
XX  
PA (TONG-) TONGHUA GANTECH BIOTECHNOLOGY LTD.  
XX  
PI Gan 2;  
XX  
DR WPI: 1999-610839/52.  
XX  
PT New chimeric proteins containing human growth hormone fragment, used particularly for the production of human insulin.



PN W09950302-A1.  
 XX 07-OCT-1999.  
 XX 31 MAR-1998: 98WO-CN000052.  
 XX 31 MAR-1998: 98WO-CN000052.  
 XX (TUNG-) TONGHUA GANTECH BIOTECHNOLOGY LTD.  
 XX Gun Z.  
 XX WPI: 1999-610839/52.  
 XX New chimeric proteins containing human growth hormone fragment, used  
 PT particularly for the production of human insulin.  
 XX Claim 14: Line 40-51: 46pp: English.  
 XX This sequence represents a chimeric protein, which contains an  
 CC N-terminal fragment of human growth hormone (hGH) of the sequence given  
 CC in AA42856, a cleavable peptide linker (AA42857), and a human insulin  
 CC precursor comprising insulin A and B chains (AA42859). The hGH portion  
 CC of the chimeric protein acts as an intramolecular chaperone (IMC) for  
 CC the insulin precursor, enabling it to fold correctly. The cleavable  
 CC peptide linker has a C-terminal Arg residue which enables the hGH portion  
 CC of the chimeric protein to be removed after folding has taken place.  
 CC Production of recombinant human insulin: an hGH-proinsulin chimeric  
 CC protein can provide human insulin with correctly linked cysteine bridges  
 CC with fewer necessary procedural steps, and hence resulting in a higher  
 CC yield of human insulin. The IMC sequences not only protect insulin  
 CC sequences from intracellular degradation by a microorganism host, but  
 CC also promote the folding of the fused insulin precursor, facilitate the  
 CC solubility of the fusion protein and decrease the intermolecular  
 CC interactions among the fusion proteins, thus allowing folding of the  
 CC fused insulin precursor at commercially useful high concentrations. The  
 CC procedural steps of cyanogen bromide cleavage, oxidative sulphitolysis  
 CC and related purification steps can thus be eliminated, along with the use  
 CC of high concentrations of mercaptan of the use of hydrophobic absorbent  
 CC resins.  
 XX  
 XX Sequence 150 AA:  
 Query Match 100.0%; Score 476; DB 20; Length 150;  
 Best Local Similarity 100.0%; Pred. No. 36-39;  
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MEFTTPLSRLEFONAMLAHRLHQLAFDTYQFEFAYIPKEQKYSFLQNPOTSLSPSESIP 60  
 DB 1 MEFTTPLSRLEFONAMLAHRLHQLAFDTYQFEFAYIPKEQKYSFLQNPOTSLSPSESIP 60  
 QY 61 TFSNREETOQKSNLELLRISILLIQLSWLEPVQ 92  
 DB 61 TFSNREETOQKSNLELLRISILLIQLSWLEPVQ 92  
 RESULT 4  
 AAF91041  
 ID AAF91041 standard; protein: 140 AA.  
 XX  
 XX AAF91041.  
 XX 14-OCT-1989 (first entry).  
 XX Human growth hormone segment.  
 XX Human growth hormone; fusion protein; thrombin;  
 KW geriatric dementia; nervous disorders; human nerve factor.  
 XX Homo sapiens (human).  
 XX EP329175-A.  
 PN

PD 23-AUG-1989.  
 XX  
 PF 17-FEB-1989; 89EP-0102795.  
 XX  
 PR 19-FEB-1988; 88JP-0035042.  
 XX  
 PA (TOXJ) TOSOH CORP.  
 XX  
 PI Ohtesuka F;  
 XX  
 XX WPI: 1989-243092/34.  
 XX  
 XX New human nerve growth factor gene encoding fusion protein  
 PT - having cleavage site for thrombin, useful for treating geriatric  
 PT dementia, etc.  
 XX  
 XX Disclosure: page 21: 48pp: English.  
 XX  
 XX Human growth hormone segment, used at the N-terminal of a fusion  
 CC protein, which contains a thrombin recognition site, and human beta nerve  
 CC growth factor (beta-NGF) at the C-terminal. Beta-NGF can be used to  
 CC control geriatric dementia and other nervous disorders, and can be  
 CC released from the fusion protein by incubation with thrombin (see  
 CC AAN90577-8, AAN91034, AAN91299).  
 XX  
 XX Sequence 140 AA:  
 Query Match 98.9%; Score 465; DB 10; Length 140;  
 Best Local Similarity 98.9%; Pred. No. 36-39;  
 Matches 91; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MEFTTPLSRLEFONAMLAHRLHQLAFDTYQFEFAYIPKEQKYSFLQNPOTSLSPSESIP 60  
 DB 1 MEFTTPLSRLEFONAMLAHRLHQLAFDTYQFEFAYIPKEQKYSFLQNPOTSLSPSESIP 60  
 QY 61 TFSNREETOQKSNLELLRISILLIQLSWLEPVQ 92  
 DB 61 TFSNREETOQKSNLELLRISILLIQLSWLEPVQ 92  
 RESULT 5  
 AAF90129  
 ID AAF90129 standard; protein: 192 AA.  
 XX  
 XX AAF90129;  
 XX  
 XX 25-MAR-2003 (updated)  
 DT 06-FEB-1996 (revised)  
 DT 01-NOV-1989 (first entry)  
 XX  
 XX Human growth hormone.  
 XX Human growth hormone; fusion protein; recombinant  
 KW vector.  
 XX  
 XX Homo sapiens (Human).  
 XX JP01144981-A.  
 XX  
 XX 07-JUN-1989.  
 XX  
 XX 02-DEC-1987; 87JP 0304937.  
 XX  
 XX 02-DEC-1987; 87JP 0304937.  
 XX  
 XX (WAKT) WAKUNAGA SEIYAKU KK.  
 PA WPI: 1989-209284/29.  
 DR N-PSDB: AAN90269.  
 XX  
 XX Recombinant vector contg. fusion protein - consisting of human  
 PT growth hormone or deriv. ligated to foreign protein, for stability  
 PT and high yield.

```

XX PS Disclosure: Fig 1: 19pp: Japanese.
XX CC
XX CC The invention consists of a vector contg. a fusion protein which is
XX CC formed by ligating, downstream of a promoter, hGH or a deriv. (pref.
XX CC formed by substn. of Met-14 with leu) and a foreign protein.
XX CC Stability of the vector in the host is greatly increased so the
XX CC protein yield is higher.
XX CC (Updated on 25-MAR-2003 to correct PA field.)
XX SO
XX Sequence 192 AA:
XX Query Match: 98.98; Score 465; DB 10; Length 192;
XX Best Local Similarity 98.98; Pred. No. 4.2e-39;
XX Matches 91; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 MPPTPLSLRFLFNAMLRHRLHQLAFDTYQFEAYIPKEQKYSFLQNPQTSLSFSSESIP 60
XX DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
XX QY 61 TPSNREETOOKSNLELLRISLLLIQSMLPEVQ 92
XX DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
XX
XX RESULT 7
XX ID AAP91299
XX XX AAP91299 standard; protein: 261 AA
XX AC AAP91299;
XX XX
XX DT 14-JEC-1989 (first entry).
XX XX
XX DE Human nerve growth factor and human growth hormone fusion protein.
XX XX
XX KW Human nerve growth factor; fusion proteins: Interleukin;
XX XX geriatric dementia; nervous disorders; human growth hormone.
XX XX
XX CS Homo sapiens (human).
XX XX
XX FH Key Location/Qualifiers
XX FT Region 1..143
XX FT Region 141..143
XX FT Region 144..161
XX XX
XX PN EP329;75-A.
XX XX
XX PD 23-AUG-1989.
XX XX
XX PF 17-FEB-1989; 89EP-0102795.
XX XX
XX PR 19-FEB-1988; 88JP-0035042.
XX XX
XX PA (TOVJ ) TOSOH CORP.
XX XX
XX PI Ohtsuka E;

```

---

```

XX PS Disclosure: Fig 1: 19pp: Japanese.
XX CC
XX CC The invention consists of a vector contg. a fusion protein which is
XX CC formed by ligating, downstream of a promoter, hGH or a deriv. (pref.
XX CC formed by substn. of Met-14 with leu) and a foreign protein.
XX CC Stability of the vector in the host is greatly increased so the
XX CC protein yield is higher.
XX CC (Updated on 25-MAR-2003 to correct PA field.)
XX SO
XX Sequence 192 AA:
XX Query Match: 98.98; Score 465; DB 10; Length 192;
XX Best Local Similarity 98.98; Pred. No. 4.2e-39;
XX Matches 91; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 MPPTPLSLRFLFNAMLRHRLHQLAFDTYQFEAYIPKEQKYSFLQNPQTSLSFSSESIP 60
XX DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
XX QY 61 TPSNREETOOKSNLELLRISLLLIQSMLPEVQ 92
XX DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
XX
XX RESULT 6
XX ID AAW92264
XX XX AAW92264 standard; Protein: 192 AA.
XX AC AAW92264;
XX XX
XX DT 08-JUN-1999 (first entry)
XX XX
XX DE Human anti-angiogenic peptide hGH Met-14pHis1.
XX XX
XX KW Human; anti-angiogenic; prolactin; placental lactogen; hPL; angiogenesis;
XX KW growth hormone; hGH; hGH-V; capillary endothelial cell proliferation;
XX KW placental vasculatization; pregnancy; treatment; angiogenic disease;
XX KW tumour; inhibitor; malignant; angiofibroma; arteriovenous malformation;
XX KW arthritis; atherosclerotic plaques; corneal graft; neovascularisation;
XX KW wound healing; proliferative retinopathy; macular degeneration; trachoma;
XX KW psoriasis; glaucoma; ocular; uveitis; fracture; Osier-Wober syndrome;
XX KW ulcers; leukaemia; scleroderma; Kaposi's sarcoma; vascular adhesion;
XX KW gene therapy; pre-eclampsia; intrauterine growth retardation;
XX KW placental dysfunction.
XX XX
XX OS Homo sapiens.
XX XX
XX PN W00851323 A1.
XX XX
XX PD 19-MAY-1998.
XX XX
XX PF 12 MAY 1998; 98WO-0509691.
XX XX
XX PR 13 MAY-1997; 97US-0046394.
XX XX
XX XX (KAGS ) UNIV CALIFORNIA.
XX PA Mortal JA, Struman L, Taylor K, Wobrow KI;
XX XX
XX PI WPI; 1999-045192/94.
XX DR N PSDB; AAX01706.
XX XX
XX XX New anti-angiogenic peptides - comprise N-terminal fragments of
XX PT human placental lactogen, human growth hormone, growth hormone
XX PT variant or human prolactin
XX XX
XX PS Example 3; Page 49; 87pp; English.
XX CC
XX CC This invention describes novel human anti-angiogenic peptides derived
XX CC from 10 to 150 consecutive amino acids selected from the N-terminal end
XX CC of human placental lactogen (hPL), human growth hormone (hGH), growth

```

```

XX WPI: 1989-0243092/34.
XX
XX New human nerve growth factor gene encoding fusion protein
PT - having cleavage site for thrombin, useful for treating geriatric
PT dementia, etc.
XX
XX Claim 36; page 31-32; 38pp; English.
XX
XX Fusion protein consisting of human growth hormone at the
CC N-terminal end (1st region), a 3 amino acid sequence representing
CC thrombin recognition site, and human beta nerve growth factor (beta-NGF)
CC at the C-terminal. Beta-NGF can be used to control geriatric dementia
CC and other nervous disorders, and can be released from the fusion
CC protein by incubation with thrombin (see AAN95577-A, AAP91034,
CC AAP91041).
XX
XX Sequence 262 AA.
XX
XX Query Match 98.9% Score 465; DB 10; Length 262;
XX Best Local Similarity 98.9% Pred. No. 49;
XX Matches 91; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 MEPTIPLSKLFDNAMLRAHRLHQLAFDTYQFEEAYIPKEQKYSFLQNPQTSLSFSESIP 60
XX
XX 1 MEPTIPLSKLFDNAMLRAHRLHQLAFDTYQFEEAYIPKEQKYSFLQNPQTSLSFSESIP 60
XX
XX 61 TPSNREETQOKSNLELLRISLLLIQSWELEPVQ 92
XX
XX 61 TPSNREETQOKSNLELLRISLLLIQSWELEPVQ 92
XX
XX
XX RESULT 8
XX AAR01740 standard; Protein: 262 AA.
XX
XX AC AAR01740;
XX
XX 25-MAR-2003 (updated)
XX 25-JUN-1991 (first entry)
XX
XX Human growth hormone/human nerve growth factor beta fusion protein.
XX hGH; hNGF; nervous system diseases; dementia.
XX
XX Homo sapiens.
XX
XX JPO1067548 A.
XX
XX 22-MAR-1991.
XX
XX 07-AUG-1989; 88JP-0202835.
XX
XX 07-AUG-1989; 88JP-0202835.
XX
XX (TOYJ) TOSOH CORP.
XX
XX WPI: 1991-128768/18.
XX N-PSDB: AAO11578.
XX
XX Purification of human neuron growth factor beta-subunit-contg. protein
PT by contacting with gel having cation exchange gp. in presence of
PT urea
XX
XX Disclosure: fig 1; 7pp; Japanese.
XX
XX A recombinant human nerve growth factor beta subunit-contg. protein
CC can be produced as this fusion protein. It is purified by contacting
CC a gel having a cation exchange gp. with the fusion protein, in the
CC presence of urea. The purified protein is useful in a medicament
CC for treating disorders of the nervous system, eg dementia.
CC (Updated on 25-MAR-2003 to correct PF field.)
XX

```

```

XX
XX Sequence 262 AA.
XX
XX Query Match 98.9% Score 465; DB 12; Length 262;
XX Best Local Similarity 98.9% Pred. No. 49;
XX Matches 91; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 MEPTIPLSKLFDNAMLRAHRLHQLAFDTYQFEEAYIPKEQKYSFLQNPQTSLSFSESIP 60
XX
XX 1 MEPTIPLSKLFDNAMLRAHRLHQLAFDTYQFEEAYIPKEQKYSFLQNPQTSLSFSESIP 60
XX
XX 61 TPSNREETQOKSNLELLRISLLLIQSWELEPVQ 92
XX
XX 61 TPSNREETQOKSNLELLRISLLLIQSWELEPVQ 92
XX
XX
XX RESULT 9
XX AAR01255 standard; Protein: 310 AA.
XX
XX AC AAR01255;
XX
XX 19-JUL-1990 (first entry)
XX
XX Fusion protein of B-cell stimulatory factor-2 and B-cell
XX differentiation factor.
XX
XX B-cell stimulatory factor-2 interleukin-6; B cell differentiation;
XX interleukin-6; fusion prote ...
XX
XX Homo sapiens.
XX
XX JPO2013375-A.
XX
XX 17-JAN-1990.
XX
XX 01-JUL-1988; 88JP-0162556.
XX
XX 01-JUL-1988; 88JP-0162556.
XX
XX (TOYJ) TOSOH CORP.
XX
XX WPI: 1990-062207/09.
XX N-PSDB: AAO02028.
XX
XX Prep'n. of human B-cell differentiation factor - from specified DNA
XX sequence segment, by recombinant DNA technique, gives protein of
XX specified amino acid sequence.
XX
XX Claim 31; Page 9; 17pp; Japanese.
XX
XX The protein is produced by fusing DNA encoding BDF (IL-) with DNA
XX encoding BDF-2 (IL-5) and ligating the product into an expression vector
XX See also AAR05311 and AAR05314.
XX
XX Sequence 310 AA;
XX
XX Query Match 98.9% Score 465; DB 11; Length 310;
XX Best Local Similarity 98.9% Pred. No. 73e-39;
XX Matches 91; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 MEPTIPLSKLFDNAMLRAHRLHQLAFDTYQFEEAYIPKEQKYSFLQNPQTSLSFSESIP 60
XX
XX 1 MEPTIPLSKLFDNAMLRAHRLHQLAFDTYQFEEAYIPKEQKYSFLQNPQTSLSFSESIP 60
XX
XX 61 TPSNREETQOKSNLELLRISLLLIQSWELEPVQ 92
XX
XX 61 TPSNREETQOKSNLELLRISLLLIQSWELEPVQ 92
XX
XX
XX RESULT 10
XX AAR05313
XX ID AAR05313 standard; Protein: 144 AA.
XX

```

```

AC AAR05313;
XX
XX DT 19-JUL-1990 (first entry)
XX
XX DE Segment of B-cell stimulatory factor-2 (IL-5).
XX
XX KW B-cell stimulatory factor-2; interleukin-5.
XX
XX OS Homo sapiens.
XX
XX PN JP02013375-A.
XX
XX PD 17-JAN-1990.
XX
XX PF 01 JUL-1988; 88JP-0162556.
XX
XX PR 01-JUL-1988; 88JP-0162556.
XX
XX PA (TOYO ) TOYO SODA MFG CO LTD.
XX
XX DR WPI: 1990 062207/09.
XX
XX DR N-PSDB: AAQ02028.
XX
XX PT Prepn. of human B-cell differentiation factor - from specified DNA
XX sequence segment, by recombinant DNA technique, gives protein of
XX specified amino acid sequence.
XX
XX PS Disclosure: Page 9; 17pp; Japanese.
XX
XX CC The sequence encoding this protein can be fused with DNA encoding B-cell
XX differentiation factor (IL-6) and ligated into an expression vector for
XX prodn. of a fusion protein.
XX
XX CC See also AAR05311.
XX
XX SQ Sequence 144 AA;
XX
XX Query Match 98.38; Score 462; DB 11; Length 144;
XX Best local Similarity 97.88; Pred. No. 1 to 39;
XX Matches 90; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 MPTTIPLSRLFDNAMLRAHRLHOLAFDTYQFEFEAYIPKEQKYSFLQNPQTSLSSESIP 60
XX
XX DB
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XX
XX QY 61 TFSNREETOQKSNLELRISILLIQSWLEPVQ 92
XX
XX DB
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XX
XX RESULT 12
XX AAP61033
XX ID AAP61033 standard; Protein: 262 AA.
XX
XX AC AAP61033;
XX
XX XX Query Match 98.38; Score 462; DB 11; Length 144;
XX DT 25-MAR-2003 (updated)
XX DT 20-NOV-1990 (first entry)
XX
XX DE Sequence of protein with somatomedin-like activity.
XX
XX KW Growth hormone.
XX
XX OS Synthetic.
XX
XX XX QY 61 TFSNREETOQKSNLELRISILLIQSWLEPVQ 92
XX
XX PN JP63167798-A.
XX
XX PD 11-JUL-1988.
XX
XX PF 29-DEC-1986; 86JP-0310177.
XX
XX PR 29-DEC 1986; 86JP-0310177.
XX
XX PA (TOYO ) TOYO SODA MFG CO LTD.
XX
XX DR WPI: 1988-232632/33.
XX
XX DR N-PSDB: AAN81605.
XX
XX PT Polypeptide with somatomedin like activity .
XX by culturing bacterium transformed by plasmid conta. gene
XX segment with specified DNA sequence.
XX
XX PS Claim 2(1); Page 609; 9pp; Japanese.
XX
XX CC The polypeptide (AAP81226) with somatomedin-like activity and the DNA
XX (AAN81605) encoding it are claimed. A Met residual gp. may be added to
XX the N-terminal. The polypeptide acts on the bone structure of mammals,
XX including humans, to promote bone growth. The polypeptide has high
XX production rate and is easily extracted from bacterial culture medium
XX and refined for use as a bone growth accelerator.
XX (updated on 25-MAR-2003 to correct PA field.)
XX

```

```

XX
XX PA (OTSU/) OTSUKA E.
XX
XX DR WPI: 1986-281696/43.
XX
XX PT Gene segment of human nerve growth factor - used in prodn. of
XX NCF-producing recombinant Escherichia strain.
XX
XX PS Claim 32; Page 482; 71pp; Japanese.
XX
XX CC The protein is a direct translation of the upstream tryptophan
XX promoter-operator lacking its attenuation sequence and human
XX beta-NGF sequence. The product may be efficiently expressed from a
XX transfected E.coli expression system.
XX
XX CC See also AAN60816-7.
XX
XX SQ Sequence 262 AA;
XX
XX Query Match 98.38; Score 462; DB 7; Length 262;
XX Best local Similarity 97.88; Pred. No. 1 to 39;
XX Matches 90; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 MEPTTIPLSRLFDNAMLRAHRLHOLAFDTYQFEFEAYIPKEQKYSFLQNPQTSLSSESIP 60
XX
XX DB
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XX
XX QY 61 TFSNREETOQKSNLELRISILLIQSWLEPVQ 92
XX
XX DB
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XX
XX RESULT 12
XX AAP81226
XX ID AAP81226 standard; protein: 138 AA.
XX
XX AC AAP81226;
XX
XX XX Query Match 98.38; Score 462; DB 7; Length 262;
XX DT 25-MAR-2003 (updated)
XX DT 20-NOV-1990 (first entry)
XX
XX DE Sequence of protein with somatomedin-like activity.
XX
XX KW Growth hormone.
XX
XX OS Synthetic.
XX
XX XX QY 61 TFSNREETOQKSNLELRISILLIQSWLEPVQ 92
XX
XX PN JP63167798-A.
XX
XX PD 11-JUL-1988.
XX
XX PF 29-DEC-1986; 86JP-0310177.
XX
XX PR 29-DEC 1986; 86JP-0310177.
XX
XX PA (TOYO ) TOYO SODA MFG CO LTD.
XX
XX DR WPI: 1988-232632/33.
XX
XX DR N-PSDB: AAN81605.
XX
XX PT Polypeptide with somatomedin like activity .
XX by culturing bacterium transformed by plasmid conta. gene
XX segment with specified DNA sequence.
XX
XX PS Claim 2(1); Page 609; 9pp; Japanese.
XX
XX CC The polypeptide (AAP81226) with somatomedin-like activity and the DNA
XX (AAN81605) encoding it are claimed. A Met residual gp. may be added to
XX the N-terminal. The polypeptide acts on the bone structure of mammals,
XX including humans, to promote bone growth. The polypeptide has high
XX production rate and is easily extracted from bacterial culture medium
XX and refined for use as a bone growth accelerator.
XX (updated on 25-MAR-2003 to correct PA field.)
XX

```

SQ Sequence 138 AA:

Query Match 97.9%; Score 460; DB 9; Length 138;  
Best Local Similarity 98.9%; Pred. No. 3e-39;  
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 FPTPLSLFVNAMLRHQLAFDTYQEFEEAYIPKEOKYSLFNPQTSLSFSESIPT 61  
|||||  
DB 1 FPTPLSLFVNAMLRHQLAFDTYQEFEEAYIPKEOKYSLFNPQTSLSFSESIPT 60  
|||||  
OY 62 PSNREETOOKSNLELLRISLLLIQSWLEPVQ 92  
|||||  
DB 61 PSNREETOOKSNLELLRISLLLIQSWLEPVQ 91  
|||||

RESULT 13

AA020110  
ID AA020110 standard: protein; 191 AA.

AC AA020110

DT 06-AUG-2002 (first entry)

DE Protein sequence of the hGH growth hormone cDNA.

KW Serum albumin-growth hormone fusion protein; growth hormone;  
KW Down's syndrome.

XX Unidentified.

OS

PN KR93C76789-A.

XX 15-OCT-1999.

XX 25-JUN-1998; 98KR-0704914.

XX 30-DEC-1995; 95GB-0026733.

PR 19-DEC-1996; 96WO-GB03164.

XX (DELZ) DELTA BIOTECHNOLOGY LTD.

XX WPI: 1997 363680/55.

DR N-PSDR; AAK99565.

XX

PT Serum albumin growth hormone fusion protein - useful to treat growth

PT hormone related diseases, e.g. Down's syndrome

PS Disclosure: Fig 1; 21pp; Korean.

XX

CC The invention relates to a serum albumin-growth hormone fusion protein -

CC useful to treat growth hormone related diseases such as Down's syndrome.

CC This sequence represents a protein of the serum albumin-growth hormone

CC cDNA of the invention.

XX

SQ Sequence 191 AA:

Query Match 97.9%; Score 460; DB 18; Length 191;

Best Local Similarity 98.9%; Pred. No. 1.3e-38;

Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 FPTPLSLFVNAMLRHQLAFDTYQEFEEAYIPKEOKYSLFNPQTSLSFSESIPT 61

|||||

DB 1 FPTPLSLFVNAMLRHQLAFDTYQEFEEAYIPKEOKYSLFNPQTSLSFSESIPT 60

|||||

OY 62 PSNREETOOKSNLELLRISLLLIQSWLEPVQ 92

|||||

DB 61 PSNREETOOKSNLELLRISLLLIQSWLEPVQ 91

|||||

RESULT 14

AA015809

ID AA015809 standard: protein; 191 AA.

XX

AC AA015809;

DT 28-JUL-1999 (first entry)

DE Primary amino acid sequence of native human growth hormone.

XX Detection; fluoresce; illegal misuse; growth substance; athlete;  
KW domesticated farm animal; cattle; human growth hormone.

XX Homo sapiens.

PN WO9926069-A1.

XX 27-MAY-1999.

XX 16-NOV-1998; 98WO-GB03449.

XX 14-NOV-1997; 97GB-0024955.

XX (GENE) GENETIC BIOLOGICALS LTD.

XX Atkinson A. Murphy JP;

XX WPI: 1999-338072/28.

XX Use of tagged exogenous polypeptide

XX Disclosure: Fig 1; 38pp; English.

XX

CC The specification describes a method of detecting an exogenously

CC administered substance from a naturally occurring endogenous substance,

CC the exogenous substance being tagged so that it fluoresces differently

CC from the endogenous one at a suitable wavelength. The tagging may

CC consist of one or more substitutions in tagged growth hormone

CC selected from G40Y, F52Y, W86F, Y, L, I or V F103Y or I137Y;

CC The method is used to distinguish between exogenously administered

CC substances as compared to naturally occurring endogenous substances.

CC Especially mentioned is the illegal misuse of growth substances by

CC athletes or in domesticated farm animals e.g. cattle. The present

CC sequence represents native human growth hormone which may be used

CC in the method of the invention.

XX

SQ Sequence 191 AA:

Query Match 97.9%; Score 460; DB 20; Length 191;

Best Local Similarity 98.9%; Pred. No. 1.3e-38;

Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 FPTPLSLFVNAMLRHQLAFDTYQEFEEAYIPKEOKYSLFNPQTSLSFSESIPT 61

|||||

DB 1 FPTPLSLFVNAMLRHQLAFDTYQEFEEAYIPKEOKYSLFNPQTSLSFSESIPT 60

|||||

OY 62 PSNREETOOKSNLELLRISLLLIQSWLEPVQ 92

|||||

DB 61 PSNREETOOKSNLELLRISLLLIQSWLEPVQ 91

|||||

RESULT 15

AA04396

ID AA04396 standard: protein; 191 AA.

XX

AC AA04396;

XX

DT 29-JUN-1999 (first entry)

XX

DE Natural human 22kDa growth hormone.

XX

KW Human; 22kDa growth hormone; hGH; mutant; thrombin; resistance;

KW plasmin; decomposition.

XX

OS Homo sapiens.

XX

PN JP11092499-A.

```
XX 06-APR-1999.
XX
XX PF 22-SEP-1997; 97JP-0275277.
XX
XX PR 22-SEP-1997; 97JP-0275277.
XX
XX PA (SUMU ) SUMITOMO SEIYAKU KK.
XX
XX DR WP1; 1999-283567/24.
XX
XX
XX PT A human growth hormone mutant - with equivalent activity to natural
XX human growth hormone
XX
XX PS Example 1: Page 5-6; 10pp; Japanese.
XX
XX CC The present invention describes a human growth hormone mutant in which
XX the 134th, Ala and the 135th Thr are replaced respectively by Asp and Pro
XX in the 1st to the 191st amino acid sequence of natural type human 22 kDa
XX growth hormone (hGH) and which has a resistance against decomposition by
XX thrombin. The present sequence represents the natural hGH. Also
XX described are: (1) a hGH mutant in which the 134th Arg, the 135th Thr
XX and the 140th Lys are replaced respectively by Asp, Pro and Ala in the
XX amino acid sequence of natural type hGH and which has a resistance
XX against decomposition by thrombin and plasmin; and (2) a drug
XX preparation containing the above hGH mutant as the active component.
XX The mutant hGH shows an activity approximately equivalent to that of
XX natural type hGH and shows a high stability in blood and body fluid.
XX
XX SQ Sequence 191 AA:
Query Match 97.9%; Score 460; DB 20; Length 191;
Best Local Similarity 98.9%; Pred. No. 1 3e-38;
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Gy 2 FPTPLSRLFDNMLRAHRLHQIAFDIYOFFEAYIPKEOKYSFTIONPOTLSLFSFSIPT 61
Db 1 FPTPLSRLFDNMLRAHRLHQIAFDIYOFFEAYIPKEOKYSFTIONPOTLSLFSFSIPT 60
Gy 62 PSNREFTQOKSNLELLKISLLILQSWLEPVQ 92
Db 61 PSNREFTQOKSNLELLKISLLILQSWLEPVQ 91
Search completed: September 15, 2003, 12:00:57
Job time : 65.6309 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:57:40 : Search time 21.1039 seconds  
(without alignments)  
184.449 Million cell updates/sec

Title: US-09-423-100-2

Perfect score: 470

Sequence: 1 MFPTPLSRLEFDNAMLRAHR.....NEELLRISILLIQSWLEPVQ 92

Scoring table: BLASTSUM62

Gapop 10.0 : Gapext 0.5

Searched: (28717 seqs, 421,058 residues)

Total number of hits satisfying chosen parameters: 32717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, AA:  
1: /cqn2\_6/plodata/1/1aa/5A.OMB.pcp:  
2: /cqn2\_6/plodata/1/1aa/5B.OMB.pcp:  
3: /cqn2\_6/plodata/1/1aa/6A.OMB.pcp:  
4: /cqn2\_6/plodata/1/1aa/6B.OMB.pcp:  
5: /cqn2\_6/plodata/1/1aa/PCITUS.OMB.pcp:  
6: /cqn2\_6/plodata/1/1aa/ba-xfiles1.pcp:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	465	98.9	192	1	Sequence 2, Appli
2	460	97.9	191	4	Sequence 5, Appli
3	460	97.9	194	2	Sequence 4, Appli
4	460	97.9	194	3	Sequence 4, Appli
5	460	97.9	217	3	Sequence 10, Appli
6	460	97.9	217	3	Sequence 10, Appli
7	460	97.9	217	3	Sequence 10, Appli
8	460	97.9	217	3	Sequence 11, Appli
9	460	97.9	217	3	Sequence 11, Appli
10	460	97.9	241	4	Sequence 25, Appli
11	460	97.9	245	4	Sequence 66, Appli
12	460	97.9	274	3	Sequence 71, Appli
13	460	97.9	360	3	Sequence 73, Appli
14	454	96.6	191	4	Sequence 1, Appli
15	454	96.6	217	1	Sequence 4, Appli
16	454	96.6	217	1	Sequence 4, Appli
17	454	96.6	217	2	Sequence 51, Appli
18	454	96.6	217	2	Sequence 51, Appli
19	454	96.6	217	4	Sequence 4, Appli
20	453	96.4	400	4	Sequence 37, Appli
21	453	96.4	401	4	Sequence 36, Appli
22	447	95.1	191	3	Sequence 18, Appli
23	445	94.7	191	3	Sequence 16, Appli
24	445	94.7	191	3	Sequence 20, Appli
25	364.5	77.6	176	3	Sequence 1, Appli
26	364.5	77.6	176	3	Sequence 1, Appli
27	358.5	76.3	176	3	Sequence 2, Appli

28 358.5 76.3 176 4 US-08-990-774-2 Sequence 2, Appli  
29 340 72.3 168 6 5424199-3 Patent No. 5424199  
30 333.5 71.0 198 1 US-08-187-756C-5 Sequence 5, Appli  
31 333.5 71.0 198 2 US-08-710-324A-5 Sequence 5, Appli  
32 333.5 71.0 198 4 US-09-411-657-5 Sequence 8, Appli  
33 306.5 65.2 191 1 US-08-468-824-8 Sequence 1, Appli  
34 304.5 64.8 191 1 US-07-963-331D-4 Sequence 4, Appli  
35 302.5 64.4 216 2 US-09-105-651-1 Sequence 1, Appli  
36 301.5 64.1 190 1 US-08-388-267C-2 Sequence 2, Appli  
37 301.5 64.1 190 4 US-09-277-720-2 Sequence 2, Appli  
38 301.5 64.1 191 6 521018C-1 Patent No. 5210180  
39 301.5 64.1 193 1 US-07-621-197C-2 Sequence 2, Appli  
40 301.5 64.1 193 1 US-08-363-982-2 Sequence 2, Appli  
41 301.5 64.1 193 2 US-08-383-621-1 Sequence 1, Appli  
42 301.5 64.1 193 3 US-08-459-906-1 Sequence 1, Appli  
43 301.5 64.1 216 2 US-09-105-651-3 Sequence 3, Appli  
44 299.5 63.7 170 1 US-07-963-331D-4 Sequence 1, Appli  
45 299.5 63.7 171 1 US-07-962-523-1 Sequence 1, Appli

## ALIGNMENTS

RESULT 1  
US-08-093-483-1  
Sequence 1, Application US/08093383  
Patent No. 5489529  
GENERAL INFORMATION:  
APPLICANT: DeBoer, Herman A.  
APPLICANT: Heyneker, Herbert L.  
APPLICANT: Seeburg, Peter H.  
TITLE OF INVENTION: DNA for Expression of Bovine Growth Hormone  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/093,483  
FILING DATE: 14-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/619827  
FILING DATE: 28-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/198824  
FILING DATE: 05-APR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 06/632361  
FILING DATE: 19-JUL-1984  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 06/303687  
FILING DATE: 18-SEP-1981  
ATTORNEY/AGENT INFORMATION:  
NAME: Johnston, Sean A.  
REGISTRATION NUMBER: P35,910  
REFERENCE/DOCKET NUMBER: 44C4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-3562  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 192 amino acids  
TYPE: amino acid

```

;
; TOPOLOGY: linear
; US-08-093-383-1
;
; Query Match 98.9%; Score 465; DB 1; Length 192;
; Best Local Similarity 98.9%; Pred. No. 1.4e-51;
; Matches 91; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
; QY 1 MFTPIPLSRFDNMLRAHRLHQLAFDTYQFFFEAYIPKQKYSFLONPOTSLSFSESIP 60
; DB 1 MFTPIPLSRFDNMLRAHRLHQLAFDTYQFFFEAYIPKQKYSFLONPOTSLSFSESIP 60
;
; QY 61 PSNREETOQKSNLELLRISLLLIQSWLEPVQ 92
; DB 61 PSNREETOQKSNLELLRISLLLIQSWLEPVQ 92
;
; RESULT 2
; US-09-284-878-5
; Sequence 5, Application US/09284878
; Patent No. 6342375
; GENERAL INFORMATION:
; APPLICANT: Olazaran, Martha Guerrero
; APPLICANT: Saldana, Hugo Barrera
; APPLICANT: Salvado, Jose Maria Vlader
; TITLE OF INVENTION: Genetically Modified
; TITLE OF INVENTION: Production and Secre
; FILE REFERENCE: 1829, 0010000
; CURRENT APPLICATION NUMBER: US/09/284, 878
; CURRENT FILING DATE: 1999-07-21
; PRIOR APPLICATION NUMBER: PCT/MX97/00043
; PRIOR FILING DATE: 1997-10-24
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 5
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-284-878-5
;
; Query Match 97.9%; Score 460; DB 4; Length 191;
; Best Local Similarity 98.9%; Pred. No. 6e-51;
; Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
; QY 2 FPTIPLSRFDNMLRAHRLHQLAFDTYQFFFEAYIPKQKYSFLONPOTSLSFSESIP 61
; DB 1 FPTIPLSRFDNMLRAHRLHQLAFDTYQFFFEAYIPKQKYSFLONPOTSLSFSESIP 60
;
; QY 62 PSNREETOQKSNLELLRISLLLIQSWLEPVQ 92
; DB 61 PSNREETOQKSNLELLRISLLLIQSWLEPVQ 91
;
; RESULT 3
; US-08-383-621-4
; Sequence 4, Application US/08383621
; Patent No. 5951972
; GENERAL INFORMATION:
; APPLICANT: Daley, Michael J.
; APPLICANT: Buckwalter, Brian L.
; APPLICANT: Gady, Susan M.
; APPLICANT: Shieh, Hong-Ming
; APPLICANT: Bohlen, Peter
; APPLICANT: Seddon, Andrew P.
; TITLE OF INVENTION: Stabilization Of Somatotropins And Other
; TITLE OF INVENTION: Proteins By Modification Of Cysteine Residues
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dr. Estelle J. Tsevdos
; STREET: 1937 West Main Street, P.O. Box 60
; CITY: Stamford
; STATE: Connecticut
; COUNTRY: U.S.A.
; ZIP: 06904-0060
;
; US-09-423-100-2
; Sequence 4, Application US/08459906
; Patent No. 6010999
; GENERAL INFORMATION:
; APPLICANT: Daley, Michael J.
; APPLICANT: Buckwalter, Brian L.
; APPLICANT: Gady, Susan M.
; APPLICANT: Shieh, Hong-Ming
; APPLICANT: Bohlen, Peter
; APPLICANT: Seddon, Andrew P.
; TITLE OF INVENTION: Stabilization Of Somatotropins And Other
; TITLE OF INVENTION: Proteins By Modification Of Cysteine Residues
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07470-8426
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/383,621
; FILING DATE: 06-FEB-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/766,142
; FILING DATE: 25-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsevdos, Estelle J.
; REGISTRATION NUMBER: 31,145
; REFERENCE/DOCKET NUMBER: 31,278-C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 203-321-2756
; TELEFAX: 203-321-2971
; TELEX: 203-710-474-4059
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 194 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-383-621-4
;
; Query Match 97.9%; Score 460; DB 2; Length 194;
; Best Local Similarity 98.9%; Pred. No. 6.1e-51;
; Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
; QY 2 FPTIPLSRFDNMLRAHRLHQLAFDTYQFFFEAYIPKQKYSFLONPOTSLSFSESIP 61
; DB 4 FPTIPLSRFDNMLRAHRLHQLAFDTYQFFFEAYIPKQKYSFLONPOTSLSFSESIP 63
;
; QY 62 PSNREETOQKSNLELLRISLLLIQSWLEPVQ 92
; DB 64 PSNREETOQKSNLELLRISLLLIQSWLEPVQ 94
;
; RESULT 4
; US-08-459-906-4
; Sequence 4, Application US/08459906
; Patent No. 6010999
; GENERAL INFORMATION:
; APPLICANT: Daley, Michael J.
; APPLICANT: Buckwalter, Brian L.
; APPLICANT: Gady, Susan M.
; APPLICANT: Shieh, Hong-Ming
; APPLICANT: Bohlen, Peter
; APPLICANT: Seddon, Andrew P.
; TITLE OF INVENTION: Stabilization Of Somatotropins And Other
; TITLE OF INVENTION: Proteins By Modification Of Cysteine Residues
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07470-8426
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,906
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Darryl L.

```

```
? REGISTRATION NUMBER: 34,276
? REFERENCE/DOCKET NUMBER: 31,278-03
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 201-831-3247
? TELEFAX: 201-831-3305
? INFORMATION FOR SEQ ID NO: 4:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 194 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
US-08-459-906-4

Query Match 97.9% Score 460; DB 3; Length 194;
Best Local Similarity 98.9% Pred. No. 7,1e-51;
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 FPTPLSRFDNAMLRAHRLHQLAFDLYOEFEAYIPKEOKYSFLONPQTSLSFSSEIPT 61
DB 4 FPTPLSRFDNAMLRAHRLHQLAFDLYOEFEAYIPKEOKYSFLONPQTSLSFSSEIPT 63

QY 62 PSNRETOOKSNLELLKRLISLLLIQSLEPVO 92
DB 64 PSNRETOOKSNLELLKRLISLLLIQSLEPVO 94

RESULT 5
US-08-589-028-10
? Sequence 10, Application US/08589028
? Patent No. 6087129
? GENERAL INFORMATION:
? APPLICANT: Newgard, Christopher B.
? APPLICANT: Halban, Philippe A.
? APPLICANT: No. 6087129mington, Karl D.
? APPLICANT: Clark, Samuel A.
? APPLICANT: Thigpen, Anice E.
? APPLICANT: Quaade, Christian
? APPLICANT: Kruse, Fred
? APPLICANT: McGarry, Dennis
? TITLE OF INVENTION: Recombinant Expression of Proteins From
? NUMBER OF SEQUENCES: 50
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Arnold, White & Durkee
? STREET: P. O. Box 4433
? CITY: Houston
? STATE: Texas
? COUNTRY: USA
? ZIP: 77210
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent In Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/589-028
? FILING DATE: Concurrently Herewith
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 60/928,427
? FILING DATE: 15-OCT-1996
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/589,028
? FILING DATE: 19-JAN-1996
? ATTORNEY/AGENT INFORMATION:
? NAME: Highland, Steven L.
? REGISTRATION NUMBER: 37,642
? REFERENCE/DOCKET NUMBER: UTSD-514
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 512/418-3000
? TELEFAX: 512/474-7577
? INFORMATION FOR SEQ ID NO: 10:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 217 amino acids
? TYPE: amino acid
? STRANDEDNESS:
? TOPOLOGY: linear
US-08-589-028-10

Query Match 97.9% Score 460; DB 3; Length 217;
Best Local Similarity 98.9% Pred. No. 7,1e-51;
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 FPTPLSRFDNAMLRAHRLHQLAFDLYOEFEAYIPKEOKYSFLONPQTSLSFSSEIPT 61
DB 27 FPTPLSRFDNAMLRAHRLHQLAFDLYOEFEAYIPKEOKYSFLONPQTSLSFSSEIPT 86

US-08-784-582-10
? Sequence 10, Application US/08784582
? Patent No. 6116707
? GENERAL INFORMATION:
? APPLICANT: Newgard, Christopher B.
? APPLICANT: Halban, Philippe A.
? APPLICANT: No. 6116707mington, Karl D.
? APPLICANT: Clark, Samuel A.
? APPLICANT: Thigpen, Anice E.
? APPLICANT: Quaade, Christian
? APPLICANT: Kruse, Fred
? APPLICANT: McGarry, Dennis
? TITLE OF INVENTION: RECOMBINANT EXPRESSION OF PROTEINS FROM
? NUMBER OF SEQUENCES: 79
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Arnold, White & Durkee
? STREET: P. O. Box 4433
? CITY: Houston
? STATE: Texas
? COUNTRY: USA
? ZIP: 77210
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent In Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/784-582
? FILING DATE: Concurrently Herewith
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 60/928,427
? FILING DATE: 15-OCT-1996
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/589,028
? FILING DATE: 19-JAN-1996
? ATTORNEY/AGENT INFORMATION:
? NAME: Highland, Steven L.
? REGISTRATION NUMBER: 37,642
? REFERENCE/DOCKET NUMBER: UTSD-514
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 512/418-3000
? TELEFAX: 512/474-7577
? INFORMATION FOR SEQ ID NO: 10:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 217 amino acids
? TYPE: amino acid
? STRANDEDNESS:
? TOPOLOGY: linear
US-08-784-582-10

Query Match 97.9% Score 460; DB 3; Length 217;
Best Local Similarity 98.9% Pred. No. 7,1e-51;
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 FPTPLSRFDNAMLRAHRLHQLAFDLYOEFEAYIPKEOKYSFLONPQTSLSFSSEIPT 61
DB 27 FPTPLSRFDNAMLRAHRLHQLAFDLYOEFEAYIPKEOKYSFLONPQTSLSFSSEIPT 86
```

OY 62 PSNRETOOKSNLELLRLISLLLIQSWLEPVQ 92  
 DB 87 PSNRETOOKSNLELLRLISLLLIQSWLEPVQ 117

## RESULT 7

US-08-785-271-10  
 : Sequence 10, Application US/08785271  
 : Patent No. 6,941,176  
 : GENERAL INFORMATION:  
 : APPLICANT: Newgard, Christopher B.  
 : APPLICANT: Halban, Philippe A.  
 : APPLICANT: No. 6194176minution, Karl D.  
 : APPLICANT: Clark, Samuel A.  
 : APPLICANT: Thigpen, Anice E.  
 : APPLICANT: Ouade, Christian  
 : APPLICANT: Kruse, Fred  
 : TITLE OF INVENTION: RECOMBINANT EXPRESSION OF PROTEINS FROM  
 : TITLE OF INVENTION: SECRETORY CELL LINES  
 : NUMBER OF SEQUENCES: 56  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Atsuro, White & Decker  
 : STREET: P.O. Box 4433  
 : CITY: Houston  
 : STATE: Texas  
 : COUNTRY: USA  
 : ZIP: 77210

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/785.271  
 FILING DATE: Concurrently herewith  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/589,028  
 FILING DATE: 19-3AN-1996

ATTORNEY/AGENT INFORMATION:  
 NAME: Highlander, Steven L.  
 REGISTRATION NUMBER: 37,642  
 REFERENCE/DOCKET NUMBER: UFSO-513  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 512/418-3000  
 TELEFAX: 512/474-7577  
 INFORMATION FOR SEQ ID NO: 10:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 217 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: linear  
 TOPOLOGY: linear

US-08-785-271-10  
 Query Match 97.9%; Score 460; DB 3; Length 217;  
 Best Local Similarity 98.9%; Pred. No. 7 1e-51;  
 Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 FPTIPLSRFDNAMLRAHRLHQLAFDTYQFFEAYIPKEOKYSLQNPOTSLSESPT 61  
 DB 27 FPTIPLSRFDNAMLRAHRLHQLAFDTYQFFEAYIPKEOKYSLQNPOTSLSESPT 86  
 OY 62 PSNRETOOKSNLELLRLISLLLIQSWLEPVQ 92  
 DB 87 PSNRETOOKSNLELLRLISLLLIQSWLEPVQ 117

## RESULT 8

US-08-759-628-11  
 : Sequence 11, Application US/08759628  
 : Patent No. 6225446  
 : GENERAL INFORMATION:  
 : APPLICANT: Altman, Scott W.

APPLICANT: Rock, Fernando L.  
 APPLICANT: Bazan, J. Fernando  
 APPLICANT: Kastelein, Robert A.  
 TITLE OF INVENTION: MUTATIONAL VARIANTS OF MAMMALIAN PROTEINS  
 NUMBER OF SEQUENCES: 11  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: DNAX Research Institute  
 STREET: 901 California Avenue  
 CITY: Palo Alto  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94304-1104  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/759.628  
 FILING DATE: 05-DEC-1996  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/008,574  
 FILING DATE: 06-DEC-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Ching, Edwin P.  
 REGISTRATION NUMBER: 34,004  
 REFERENCE/DOCKET NUMBER: 1X05520  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-852-9196  
 TELEFAX: 415-496-1400  
 INFORMATION FOR SEQ ID NO: 11:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 217 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 FEATURE:  
 NAME/KEY: Peptide  
 LOCATION: 32..53  
 FEATURE:  
 NAME/KEY: Peptide  
 LOCATION: 94..115  
 FEATURE:  
 NAME/KEY: Peptide  
 LOCATION: 133..153  
 FEATURE:  
 NAME/KEY: Peptide  
 LOCATION: 192..210  
 OTHER INFORMATION: /note: "The peptides above are  
 : OTHER INFORMATION: depicted in Figure 1"  
 US-08-759-628-11

Query Match 97.9%; Score 460; DB 3; Length 217;  
 Best Local Similarity 98.9%; Pred. No. 7 1e-51;  
 Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 FPTIPLSRFDNAMLRAHRLHQLAFDTYQFFEAYIPKEOKYSLQNPOTSLSESPT 61  
 DB 27 FPTIPLSRFDNAMLRAHRLHQLAFDTYQFFEAYIPKEOKYSLQNPOTSLSESPT 86  
 OY 62 PSNRETOOKSNLELLRLISLLLIQSWLEPVQ 92  
 DB 87 PSNRETOOKSNLELLRLISLLLIQSWLEPVQ 117

## RESULT 9

US-09-284-878-1  
 : Sequence 1, Application US/09284878  
 : Patent No. 6342375  
 : GENERAL INFORMATION:  
 : APPLICANT: Olazaran, Martha Guerrero

APPLICANT: Saldana, Hugo Barrera  
TITLE OF INVENTION: Genetically Modified Methylothrophic P. pastoris Yeast for the  
TITLE OF INVENTION: Production and Secretion of the Human Growth Hormone  
FILE REFERENCE: 1829.0010000  
CURRENT APPLICATION NUMBER: US/09/284,878  
PRIOR FILING DATE: 1999-07-21  
PRIOR APPLICATION NUMBER: PCT/MX97/00033  
PRIOR FILING DATE: 1997-10-24  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 217  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-284,878-1

Query Match: 97.9%; Score 460; DB 4; Length 241;  
Best Local Similarity 98.9%; Pred. No. 8.2e-51;  
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 FPTPLSRFDNAMLRAHRLHQLAFDTYQFEAYIPKEOKYSFLONPOTSLSFSES IPT 61  
|||||  
Db 27 FPTPLSRFDNAMLRAHRLHQLAFDTYQFEAYIPKEOKYSFLONPOTSLSFSES IPT 86  
|||||

QY 62 PSNRETOOKSNLELLRISLLLIQSWLEPVQ 92  
|||||  
Db 87 PSNRETOOKSNLELLRISLLLIQSWLEPVQ 117  
|||||

RESULT 10  
US-09-424,620B-25  
Sequence 25, Application US/09424620B  
Patent No. 6391585  
GENERAL INFORMATION:  
APPLICANT: HANIL SYNTHETIC FIBER CO., LTD.  
JANG, Ki-Ryong  
MOON, Jae-Woong  
HAE, Cheon-Soon  
YANG, Doo-Suk  
LEE, Jee-Won  
SEONG, Baik-Lin  
TITLE OF INVENTION: Process for preparing recombinant proteins using highly  
efficient expression vector from Saccharomyces cerevisiae  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BACHMAN & LAPORTE, P.C.  
STREET: Suite 1201, 900 Chapel Street  
CITY: New Haven  
STATE: Connecticut  
COUNTRY: U.S.A.  
ZIP: 06510-2802  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
COMPUTER: IBM  
OPERATING SYSTEM: WINDOWS 95/98  
SOFTWARE: MS WORD  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/424,620B  
FILING DATE: 24-Nov-03 6391585 1999  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 241 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: PROTEIN  
SEQUENCE DESCRIPTION: SEQ ID NO: 25:  
US-09-424,620B-25

Query Match: 97.9%; Score 460; DB 4; Length 241;  
Best Local Similarity 98.9%; Pred. No. 8.2e-51;  
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 FPTPLSRFDNAMLRAHRLHQLAFDTYQFEAYIPKEOKYSFLONPOTSLSFSES IPT 61  
|||||  
Db 51 FPTPLSRFDNAMLRAHRLHQLAFDTYQFEAYIPKEOKYSFLONPOTSLSFSES IPT 110  
|||||

QY 62 PSNRETOOKSNLELLRISLLLIQSWLEPVQ 92  
|||||  
Db 111 PSNRETOOKSNLELLRISLLLIQSWLEPVQ 141  
|||||

RESULT 11  
US-09-280,030-66  
Sequence 66, Application US/09280030A  
Patent No. 6506595  
GENERAL INFORMATION:  
APPLICANT: Sato, Seiji  
APPLICANT: HIKASHIKAMI, Naotaka  
APPLICANT: Kudo, Toshiyuki  
TITLE OF INVENTION: DNAs encoding new fusion proteins and processes for  
TITLE OF INVENTION: PREPARING USEFUL POLYPEPTIDES THROUGH EXPRESSION OF THE  
TITLE OF INVENTION: DNAs  
FILE REFERENCE: 382,1026  
CURRENT APPLICATION NUMBER: US/09/280,030A  
CURRENT FILING DATE: 1999-03-26  
EARLIER APPLICATION NUMBER: JP10-87339/1998  
EARLIER FILING DATE: 1998-03-31  
NUMBER OF SEQ ID NOS: 66  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 66  
LENGTH: 245  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Designated is  
OTHER INFORMATION: an amino acid sequence of MDPsp-MWmp20 TEV-G-GH  
US-09-280,030-66

Query Match: 97.9%; Score 460; DB 4; Length 245;  
Best Local Similarity 98.9%; Pred. No. 8.4e-51;  
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 FPTPLSRFDNAMLRAHRLHQLAFDTYQFEAYIPKEOKYSFLONPOTSLSFSES IPT 61  
|||||  
Db 55 FPTPLSRFDNAMLRAHRLHQLAFDTYQFEAYIPKEOKYSFLONPOTSLSFSES IPT 114  
|||||

QY 62 PSNRETOOKSNLELLRISLLLIQSWLEPVQ 92  
|||||  
Db 115 PSNRETOOKSNLELLRISLLLIQSWLEPVQ 145  
|||||

RESULT 12  
US-08-784,592-71  
Sequence 71, Application US/08784592  
Patent No. 6110707  
GENERAL INFORMATION:  
APPLICANT: Newgard, Christopher B.  
APPLICANT: Halban, Philippe A.  
APPLICANT: No. 6110707minington, Karl D.  
APPLICANT: Clark, Samuel A.  
APPLICANT: Thigpen, Anice E.  
APPLICANT: Ouade, Christian  
APPLICANT: Kruse, Fred  
APPLICANT: McGarry, Dennis  
TITLE OF INVENTION: RECOMBINANT EXPRESSION OF PROTEINS FROM  
TITLE OF INVENTION: SECRETORY CELL LINES  
NUMBER OF SEQUENCES: 79  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210



```

1 Patent No. 5897709
2 GENERAL INFORMATION:
3 APPLICANT: ROSEN, ET AL.
4 TITLE OF INVENTION: Human Growth Hormone
5 NUMBER OF SEQUENCES: 7
6 CORRESPONDENCE ADDRESS:
7 ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLIAN,
8 ADDRESSEE: CECCHI, STEWART & OLSTEIN
9 STREET: 6 BECKER FARM ROAD
10 CITY: ROSELAND
11 STATE: NEW JERSEY
12 COUNTRY: USA
13 ZIP: 07068
14 COMPUTER READABLE FORM:
15 MEDIUM TYPE: 3.5 INCH DISKETTE
16 COMPUTER: IBM PS/2
17 OPERATING SYSTEM: MS-DOS
18 SOFTWARE: WORD PERFECT 5.1
19 CURRENT APPLICATION DATA:
20 APPLICATION NUMBER: US/08/187,709
21 FILING DATE: January 27, 1994
22 CLASSIFICATION: 435
23 PRIOR APPLICATION DATA:
24 APPLICATION NUMBER:
25 FILING DATE:
26 ATTORNEY/AGENT INFORMATION:
27 NAME: FERRARO, GREGORY D.
28 REGISTRATION NUMBER: 36,134
29 REFERENCE/DOCKET NUMBER: 325820-55
30 TELECOMMUNICATION INFORMATION:
31 TELEPHONE: 201-994-1700
32 TELEFAX: 201-994-1744
33 INFORMATION FOR SEQ ID NO: 4:
34 SEQUENCE CHARACTERISTICS:
35 LENGTH: 217 AMINO ACIDS
36 TYPE: AMINO ACID
37 STRANDEDNESS:
38 TOPOLOGY: LINEAR
39 MOLECULE TYPE: PROTEIN
40 US-08-187-756C-4

```

```

Query Match 96.6% Score 454; DB 1; Length 217;
Best Local Similarity 97.8% Pred No. 4, 2e-50;
Matches 89; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 FTPTPLSRFDNAMLRAHRLHQLAFDYOEFEAYIPKEQKYSFLQNPQTSLSFSES IPT 61
DB 27 FTPTPLSRFDNAMLRAHRLHQLAFDYOEFEAYIPKEQKYSFLQNPQTSLSFSES IPT 86
QY 62 PSNREETOOKSNLELLRLISLLIQSMLEPVQ 92
DB 87 PSNREETOOKSNLELLRLISLLIQSMLEPVQ 117

```

Search completed: September 15, 2003, 12:05:31  
Job time : 22.1039 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: September 15, 2003, 12:04:45 : Search time 38.5606 seconds  
(without alignments)  
347.945 Million cell updates/sec

Title: US-09-423-100-2

Perfect score: 470

Sequence: 1 MEPTIPLSRLFDNMLKAHR.....NLEELRISLLILOSILWLPVQ 42

Scoring table: HIGSOM62

Gapop 10.0, Gapext 0.5

Searched: 54,946 seqs, 1459,246 residues

Total number of hits satisfying chosen parameters: 4

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA:\*

1: /cqn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cqn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cqn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cqn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cqn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cqn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cqn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cqn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cqn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cqn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cqn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cqn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cqn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cqn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cqn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cqn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cqn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cqn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	470	100.0	92	14	US-10-054-8 ( 2
2	470	100.0	150	14	US-10-054-8.3-7
3	460	97.9	191	11	US-09-984-010-23
4	460	97.9	191	12	US-10-153-207-1
5	460	97.9	191	12	US-10-400-177-1
6	460	97.9	191	12	US-10-400-708-1
7	460	97.9	214	12	US-10-153-207-6
8	460	97.9	215	9	US-09-929-918-9
9	460	97.9	247	9	US-09-280-030-66
10	454	96.6	217	9	US-09-853-688-2
11	454	96.6	217	12	US-09-969-748C-4
12	453	96.4	217	10	US-09-804-409A-16
13	447	95.1	217	9	US-09-853-688-4
14	444	94.5	191	12	US-09-824-200-12
15	396	84.3	217	9	US-09-850-887-3

16	381	81.1	163	15	US-10-043-487-350
17	381	81.1	191	12	US-10-153-207-2
18	373	79.4	229	15	US-10-103-313-411
19	345	73.4	246	15	US-10-188-246-18
20	301.5	64.1	190	12	US-10-153-207-3
21	260	55.3	49	14	US-10-054-873-1
22	260	55.3	107	14	US-10-054-873-6
23	174.5	37.1	188	15	US-10-191-879-19
24	174.5	37.1	210	15	US-10-191-879-10
25	162	34.5	136	15	US-10-043-487-337
26	161.5	34.4	386	10	US-09-887-569A-2
27	151.5	34.4	388	12	US-10-122-746-4
28	142	30.2	54	9	US-09-876-478-14
29	131	27.9	198	15	US-10-140-293-32
30	126	26.8	199	15	US-10-140-293-33
31	125	26.6	199	15	US-10-140-293-26
32	125	26.6	199	15	US-10-140-293-27
33	124	25.4	199	15	US-10-140-293-31
34	123	26.2	199	15	US-10-140-293-22
35	122	26.0	199	15	US-10-140-293-21
36	121.5	25.9	124	14	US-10-046-869-26
37	121.5	25.9	208	15	US-10-191-879-2
38	121.5	25.9	253	14	US-10-036-869-27
39	119	25.3	199	15	US-10-140-293-24
40	118	25.1	199	15	US-10-140-293-18
41	118	25.1	199	15	US-10-140-293-19
42	117	24.9	197	15	US-10-140-293-15
43	116	24.7	197	15	US-10-140-293-13
44	116	24.7	199	12	US-10-153-207-4
45	116	24.7	199	15	US-10-140-293-20

#### ALIGNMENTS

#### RESULT 1

US-10-054-873-2

: Sequence 2, Application US/10054873

: Publication No. US20020164712A1

: GENERAL INFORMATION:

: APPLICANT: Gan, Zhong Ru

: TITLE OF INVENTION: Chimeric Protein Containing an

: Intramolecular Chaperone-like Sequence

: NUMBER OF SEQUENCES: 7

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Townsend and Townsend and Crew LLP

: STREET: Two Embarcadero Center, Eighth Floor

: CITY: San Francisco

: STATE: California

: COUNTRY: USA

: ZIP: 94111-4334

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: PatentIn Release #1.0, Version #1.30

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/10/054-873

: FILING DATE: 22-Jan-2002

: CLASSIFICATION: <Unknown>

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: WO PCT/CN98/00052

: FILING DATE: 31-MAR-1998

: APPLICATION NUMBER: US 09/423,100

: FILING DATE: 11-DEC-2000

: ATTORNEY/AGENT INFORMATION:

: NAME: Mycroft, Frank J

: REGISTRATION NUMBER: 46,946

: REFERENCE/DOCKET NUMBER: 020167-000130US

: INFORMATION FOR SEQ ID NO: 2:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 92 amino acids

: TYPE: amino acid



2 PRIOR FILING DATE: 1992-10-13  
2 PRIOR APPLICATION NUMBER: 07/875,204  
2 PRIOR FILING DATE: 1992-04-27  
2 PRIOR APPLICATION NUMBER: 07/428,066  
2 PRIOR FILING DATE: 1989-10-26  
2 PRIOR APPLICATION NUMBER: 07/264,611  
2 PRIOR FILING DATE: 1988-10-28  
2 NUMBER OF SEQ ID NOS: 20  
2 SOFTWARE: FastSeq for Windows Version 4.0  
2 SEQ ID NO 1  
2 LENGTH: 191  
2 TYPE: PRT  
2 ORGANISM: Homo Sapiens  
US-10-153-207-1

Query Match 97.9% Score 460; DB 12; Length 191;  
Best Local Similarity 98.9% Pred. No. 78e-45;  
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 FPTPLSKLFONAMLAHRLHQLAFDTYOFEEAYIPKEQKYSFLONPQTSLSSESPT 61  
|||||  
Db 1 FPTPLSKLFONAMLAHRLHQLAFDTYOFEEAYIPKEQKYSFLONPQTSLSSESPT 60  
QY 62 PSNRETOOKSNLELLRLISLLLIQSWLEPVQ 92  
|||||  
Db 61 PSNRETOOKSNLELLRLISLLLIQSWLEPVQ 91

RESULT 5  
US-10-400-377-1  
2 Sequence 1, Application US/10400377  
2 Publication No. US20030162949A1  
2 GENERAL INFORMATION:  
2 APPLICANT: Cox III, George N  
2 TITLE OF INVENTION: Derivatives of Growth Hormone and Related Proteins  
2 FILE REFERENCE: 4152-1-PUS  
2 CURRENT APPLICATION NUMBER: US/10/400,377  
2 CURRENT FILING DATE: 2003-03-26  
2 PRIOR APPLICATION NUMBER: US/09/462,941  
2 PRIOR FILING DATE: 2000-01-14  
2 PRIOR APPLICATION NUMBER: 60/052,516  
2 PRIOR FILING DATE: 1997-07-14  
2 NUMBER OF SEQ ID NOS: 41  
2 SOFTWARE: Patent In Ver. 2.0  
2 SEQ ID NO 1  
2 LENGTH: 191  
2 TYPE: PRT  
2 ORGANISM: Homo Sapiens  
US-10-400-377-1

Query Match 97.9% Score 460; DB 12; Length 191;  
Best Local Similarity 98.9% Pred. No. 78e-45;  
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 FPTPLSKLFONAMLAHRLHQLAFDTYOFEEAYIPKEQKYSFLONPQTSLSSESPT 61  
|||||  
Db 1 FPTPLSKLFONAMLAHRLHQLAFDTYOFEEAYIPKEQKYSFLONPQTSLSSESPT 60  
QY 62 PSNRETOOKSNLELLRLISLLLIQSWLEPVQ 92  
|||||  
Db 61 PSNRETOOKSNLELLRLISLLLIQSWLEPVQ 91

RESULT 6  
US-10-400-798-1  
2 Sequence 1, Application US/10400708  
2 Publication No. US20030166865A1  
2 GENERAL INFORMATION:  
2 APPLICANT: Cox III, George N  
2 TITLE OF INVENTION: Derivatives of Growth Hormone and Related Proteins  
2 FILE REFERENCE: 4152-1-PUS

2 CURRENT APPLICATION NUMBER: US/10/400,798  
2 CURRENT FILING DATE: 2003-03-26  
2 PRIOR APPLICATION NUMBER: US/09/462,941  
2 PRIOR FILING DATE: 2000-01-14  
2 PRIOR APPLICATION NUMBER: 60/052,516  
2 PRIOR FILING DATE: 1997-07-14  
2 NUMBER OF SEQ ID NOS: 41  
2 SOFTWARE: Patent In Ver. 2.0  
2 SEQ ID NO 1  
2 LENGTH: 191  
2 TYPE: PRT  
2 ORGANISM: Homo Sapiens  
US-10-400-798-1

Query Match 97.9% Score 460; DB 12; Length 191;  
Best Local Similarity 98.9% Pred. No. 78e-45;  
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 FPTPLSKLFONAMLAHRLHQLAFDTYOFEEAYIPKEQKYSFLONPQTSLSSESPT 61  
|||||  
Db 1 FPTPLSKLFONAMLAHRLHQLAFDTYOFEEAYIPKEQKYSFLONPQTSLSSESPT 60  
QY 62 PSNRETOOKSNLELLRLISLLLIQSWLEPVQ 92  
|||||  
Db 61 PSNRETOOKSNLELLRLISLLLIQSWLEPVQ 91

RESULT 7  
US-10-153-207-6  
2 Sequence 6, Application US/10153207  
2 Publication No. US20030153003A1  
2 GENERAL INFORMATION:  
2 APPLICANT: James A. Wells  
2 TITLE OF INVENTION: GROWTH HORMONE VARIANTS  
2 FILE REFERENCE: 669,12-US-C7  
2 CURRENT APPLICATION NUMBER: US/10/153,207  
2 CURRENT FILING DATE: 2002-05-22  
2 PRIOR APPLICATION NUMBER: 08/479,884  
2 PRIOR FILING DATE: 1995-06-07  
2 PRIOR APPLICATION NUMBER: 08/190,723  
2 PRIOR FILING DATE: 1994-02-02  
2 PRIOR APPLICATION NUMBER: 07/960,227  
2 PRIOR FILING DATE: 1992-10-13  
2 PRIOR APPLICATION NUMBER: 07/875,204  
2 PRIOR FILING DATE: 1992-04-27  
2 PRIOR APPLICATION NUMBER: 07/428,066  
2 PRIOR FILING DATE: 1989-10-26  
2 PRIOR APPLICATION NUMBER: 07/264,611  
2 PRIOR FILING DATE: 1988-10-28  
2 NUMBER OF SEQ ID NOS: 20  
2 SOFTWARE: FastSeq for Windows Version 4.0  
2 SEQ ID NO 6  
2 LENGTH: 214  
2 TYPE: PRT  
2 ORGANISM: Homo Sapiens  
US-10-153-207-6

Query Match 97.9% Score 460; DB 12; Length 214;  
Best Local Similarity 98.9% Pred. No. 9e-45;  
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 FPTPLSKLFONAMLAHRLHQLAFDTYOFEEAYIPKEQKYSFLONPQTSLSSESPT 61  
|||||  
Db 24 FPTPLSKLFONAMLAHRLHQLAFDTYOFEEAYIPKEQKYSFLONPQTSLSSESPT 83  
QY 62 PSNRETOOKSNLELLRLISLLLIQSWLEPVQ 92  
|||||  
Db 84 PSNRETOOKSNLELLRLISLLLIQSWLEPVQ 114

RESULT 8  
US-09-929-918-9

```

RESULT 11
US-09-969-748C-4
: Sequence 4, Application US/09969748C
: Publication No. US20030161809A1
: GENERAL INFORMATION:
: APPLICANT: ARIZEKE PHARMACEUTICALS, INC.
: APPLICANT: HOUSTON, Lou, L.
: APPLICANT: SHERIDAN, Philip, J.
: APPLICANT: HAWLEY, Stephen
: APPLICANT: GLYNN, Jacqueline, M.
: APPLICANT: CHAPIN, Steven
: APPLICANT: BASU, Amarendra
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TRANSPORT OF BIOLOGICALLY AC
: TITLE OF INVENTION: AGENTS ACROSS CELLULAR BARRIERS
: FILE REFERENCE: 057220-0303
: CURRENT APPLICATION NUMBER: US/09/969,748C
: CURRENT FILING DATE: 2002-12-10
: PRIOR APPLICATION NUMBER: US 60/267,601
: PRIOR FILING DATE: 2001-02-09
: PRIOR APPLICATION NUMBER: US 60/248,819
: PRIOR FILING DATE: 2000-11-14
: PRIOR APPLICATION NUMBER: US 60/248,478
: PRIOR FILING DATE: 2000-11-13
: PRIOR APPLICATION NUMBER: US 60/217,929
: PRIOR FILING DATE: 2000-10-02
: NUMBER OF SEQ ID NOS: 115
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 4
: LENGTH: 217
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-969-748C-4
Query Match          96.6%   Score 454;   DB 12;
Best Local Similarity 97.8%   Pred. No. 4.5e-44;
Length 217;

```

```

Matches 89; Conservative 0; Mismatch: 5 2; Indels 0; Gaps 0;

QY 2 FPTIPLSRFDNMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLONPOTSLSFSES IPT 61
DB 27 FPTIPLSRFDNMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLONPOTSLSFSES IPT 86

QY 62 PSNREETOOKSNLELRISLLLIQSMLRPVQ 92
DB 87 PSNREETOOKSNLELRISLLLIQSMLRPVQ 117

RESULT 12
US-09-804-409A-16
; Sequence 16, Application US/09804409A
; Patent No. US2002015100A1
; GENERAL INFORMATION:
; APPLICANT: KIEFER, TIMOTHY J.
; APPLICANT: CHEN, ANTHONY T.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATED PROTEIN
; FILE REFERENCE: 025996/027 8721
; CURRENT APPLICATION NUMBER: US/09/804,409A
; CURRENT FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 16
; LENGTH: 217
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-804-409A-16

Query Match 96.4%; Score 453; DB 10; Length 217;
Best Local Similarity 97.8%; Pred. No. 5, 30-43;
Matches 89; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 FPTIPLSRFDNMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLONPOTSLSFSES IPT 61
DB 27 FPTIPLSRFDNMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLONPOTSLSFSES IPT 86

QY 62 PSNREETOOKSNLELRISLLLIQSMLRPVQ 92
DB 87 PSNREETOOKSNLELRISLLLIQSMLRPVQ 117

RESULT 13
US-09-853-688-4
; Sequence 4, Application US/09853688
; Patent No. US20020081605A1
; GENERAL INFORMATION:
; APPLICANT: COOPER, DAVID N.
; APPLICANT: PROCTOR, ANNIE M.
; APPLICANT: GREGORY, JOHN
; APPLICANT: MILLAR, DAVID S.
; TITLE OF INVENTION: METHOD FOR DETECTING GROWTH HORMONE VARIATIONS IN
; FILE REFERENCE: WCM78
; CURRENT APPLICATION NUMBER: US/09/853,688
; CURRENT FILING DATE: 2001-05-14
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 4
; LENGTH: 217
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-853-688-4

Query Match 95.1%; Score 447; DB 9; Length 217;
Best Local Similarity 96.7%; Pred. No. 1, 8e-43;
Matches 88; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 FPTIPLSRFDNMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLONPOTSLSFSES IPT 61
DB 27 FPTIPLSRFDNMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLONPOTSLSFSES IPT 86

```

```

QY 62 PSNREETOOKSNLELRISLLLIQSMLRPVQ 92
DB 87 PSNREETOOKSNLELRISLLLIQSMLRPVQ 117

RESULT 14
US-09-824-200-12
; Sequence 12, Application US/09824200
; Publication No. US20030167531A1
; GENERAL INFORMATION:
; APPLICANT: RUSSELL, DOUGLAS A.
; APPLICANT: SCHLITTLER, MICHAEL
; TITLE OF INVENTION: EXPRESSION AND PURIFICATION OF BIOACTIVE, AUTHENTIC
; FILE REFERENCE: 16712.003;
; CURRENT APPLICATION NUMBER: US/09/824,200
; PRIOR FILING DATE: 2001-04-03
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-824-200-12

Query Match 94.5%; Score 444; DB 12; Length 191;
Best Local Similarity 96.7%; Pred. No. 5, 30-43;
Matches 88; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 FPTIPLSRFDNMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLONPOTSLSFSES IPT 61
DB 1 FPTIPLSRFDNMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLONPOTSLSFSES IPT 60

QY 62 PSNREETOOKSNLELRISLLLIQSMLRPVQ 92
DB 61 PSNREETOOKSNLELRISLLLIQSMLRPVQ 91

RESULT 15
US-09-850-887-3
; Sequence 3, Application US/09850887
; Patent No. US20020009778A1
; GENERAL INFORMATION:
; APPLICANT: Gal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Gargano, Gina
; TITLE OF INVENTION: THYROID AND MITOCHONDRIAL MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/850,887
; FILING DATE: 07 May 2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/087,678
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:56:45 ; Search time 15.1685 seconds  
(without alignments)  
583,284 Million cell updates/sec

Title: US-09-423-100-2

Perfect score: 470

Sequence: 1 MFPTPLSRLEFONAMLAHR.....NLELLRSLILQSWLEPVQ 92

Scoring table: BLOSUM62

Gapop 10 0 , Gapext 0 0

Searchset 25000 seqs, 60158682 test seqs

Total number of hits satisfying chosen parameters: 283408

Minimum DB seq length: 5

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	460	97.9	217	1	STHO
2	460	97.9	217	2	somatotropin 1 pre
3	422	89.8	217	1	STHOV2
4	422	89.8	217	2	somatotropin 2 pre
5	402	85.5	217	2	167411
6	397	84.5	217	2	167409
7	396	84.3	212	2	167408
8	396	84.3	217	2	153267
9	381	81.1	217	1	LCRUC
10	381	81.1	217	2	E32435
11	359.5	76.5	215	2	A26449
12	310.5	66.1	216	2	B49159
13	307.5	65.4	190	2	PN0140
14	304.5	64.8	216	1	STMS
15	302.5	64.4	190	1	STHO
16	302.5	64.4	216	1	STRT
17	302.5	64.4	216	2	S49481
18	301.5	64.1	190	2	JKC219
19	301.5	64.1	216	1	STPG
20	301.5	64.1	216	2	I46145
21	301.5	64.1	216	2	JC4632
22	299.5	63.7	216	2	A37782
23	297.5	63.3	190	1	A61584
24	295.5	62.9	190	2	JS0429
25	289.5	61.6	217	1	STBO
26	289.5	61.6	217	1	STSH
27	289.5	61.6	217	1	STGT
28	289.5	61.6	217	2	S32682
29	278.5	59.3	216	2	JC1514

30	275.5	58.6	216	2	A60509
31	268.5	57.1	191	2	A60625
32	261	55.5	216	2	S04929
33	257.5	54.8	190	2	S21750
34	247.5	52.7	190	2	A56816
35	238.5	50.7	215	2	151188
36	237.5	50.5	215	2	JS0037
37	234	49.8	199	2	B32435
38	233.5	49.7	195	2	151250
39	225.5	48.0	183	2	A60623
40	206	43.8	87	4	167761
41	174.5	37.1	209	2	JT0483
42	171	36.4	163	2	JN0387
43	165.5	35.2	190	2	JC5682
44	165.5	35.2	210	2	150763
45	165.5	35.2	210	2	S36351

ALIGNMENTS

## RESULT 1

STHO

somatotropin 1 precursor [validated] - human

N:Alternate names: growth hormone 1; GH-N; pituitary somatotropin

N:Contains: growth hormone 5K peptide; somatotropin 1, long form; somatotropin 1, s

C:Species: Homo sapiens (man)

C:Date: 24-Apr-1984 #sequence revision 10-Feb-1995 #text change 08-Dec-2000

C:Accession: A93731; A32435; A93694; A94247; A90651; A93357; A91778; A91764; A90217

R:DeNoto, F.M.; Moore, B.D.; Goodman, H.M.

Nucleic Acids Res. 9, 3719-3730, 1981

A:Title: Human growth hormone DNA sequence and mRNA structure: possible alternative

A:Reference number: A93731; MUID:82014939; PMID:6259091

A:Accession: A93731

A:Molecule type: DNA

A:Residues: 1-217 <DEN>

A:Cross-references: GB:V00520

A>Note: the 20K short form somatotropin lacks residues 58-72 (32-46 in the active h

R:Chen, E.Y.; Liao, Y.C.; Smith, D.H.; Barrera-Saldana, H.A.; Gellinas, R.E.; Seebur

Genomics 4, 479-497, 1989

A:Title: The human growth hormone locus: nucleotide sequence, biology, and evolutio

A:Reference number: A32435; MUID:89307277; PMID:2744760

A:Accession: A32435

A:Molecule type: DNA

A:Residues: 1-217 <CHE>

A:Cross-references: GB:V00071; MUID:q183148; PIDN:AAA52549.i; PMID:183349

R:Koskam, W.; Rougeon, F.

Nucleic Acids Res. 7, 305-320, 1979

A:Title: Molecular cloning and nucleotide sequence of the human growth hormone stru

A:Reference number: A93694; MUID:80034477; PMID:686281

A:Accession: A93694

A:Molecule type: mRNA

A:Residues: 1-217 <ROS>

A:Cross-references: GB:V00519

A>Note: 35-Pro was also found

R:Martial, J.A.; Hallewell, R.A.; Baxter, J.D.; Goodman, H.M.

Science 205, 602-607, 1979

A:Title: Human growth hormone: complementary DNA cloning and expression in bacteria

A:Reference number: A94247; MUID:79203293; PMID:477496

A:Accession: A94247

A:Molecule type: mRNA

A:Residues: 1-217 <MAR>

R:Li, C.H.; Dixon, J.S.; Liu, W.K.

Arch. Biochem. Biophys. 133, 70-9, 1969

A:Title: Human pituitary growth hormone. XIX. The primary structure of the hormone.

A:Reference number: A90048; MUID: 9289202; PMID:5810834

A:Contents: annotation

R:Li, C.H.; Dixon, J.S.

Arch. Biochem. Biophys. 146, 233-236, 1971

A:Title: Human pituitary growth hormone. XXXII. The primary structure of the hormon

A:Reference number: A90051; MUID:72143935; PMID:5144027

A:Accession: A90051

A:Molecule type: protein

A:Residues: 27-94;96-217 <LIC>  
 R:Niall, H.D.  
 Nature New Biol. 230, 90-91, 1971  
 A:Title: Revised primary structure for human growth hormone.  
 A:Reference number: A93397; MUID:71139765; PMID:5275046  
 A:Accession: A93397  
 A:Molecule type: protein  
 A:Residues: 27-51 <NIA>  
 R:Niall, H.D.; Hogan, M.L.; Sauer, R.; Rosenblum, I.Y.; Greenwood, F.C.  
 Proc. Natl. Acad. Sci. U.S.A. 68, 866-869, 1971  
 A:Title: Sequences of pituitary and placental lactogenic and growth hormones: evolution  
 A:Reference number: A93778; MUID:71153968; PMID:5275528  
 A:Accession: A93778  
 A:Molecule type: protein  
 A:Residues: 119-120;157-159 <N12>  
 R:Niall, H.D.  
 In Prolactin and Carcinogenesis, Paper Fourth Terasus Workshop, Proceedings, K. J. Lewis, G. J. Jonowald, L.F.; Lewis, L.J.  
 Biochem. Biophys. Res. Commun. 92, 511-516, 1980  
 A:Title: The 20,000-dalton variant of human growth hormone: location of the amino acid  
 A:Reference number: A90217; MUID:80130196; PMID:7356479  
 A:Accession: A90217  
 A:Contents: somatotropin, 20K short variant  
 A:Title: A90217  
 A:Molecule type: protein  
 A:Residues: 46-57;73-80 <LEW>  
 R:Chapman, G.E.; Rogers, K.M.; Brittain, L.; Bradshaw, R.A.; Bates, G.J.; Turner, C.; Ca  
 J. Biol. Chem. 256, 2395-2401, 1981  
 A:Title: The 20,000 molecular weight variant of human growth hormone. Preparation and so  
 A:Reference number: A92311; MUID:81117361; PMID:7462247  
 A:Accession: A92311  
 A:Contents: somatotropin, 20K short variant  
 A:Title: A92311  
 A:Molecule type: protein  
 A:Residues: 27-57;73-79 <CHA>  
 R:Singh, R.N.P.; Shavey, B.K.; Lewis, L.J.; Lewis, U.J.  
 J. Protein Chem. 2, 425-436, 1983  
 A:Title: Human growth hormone peptide 1-43: isolation from pituitary glands.  
 A:Reference number: A61466  
 A:Accession: A61466  
 A:Molecule type: protein  
 A:Residues: 27-69 <SIN>  
 R:Robson, V.M.J.; Rae, I.D.; NG, F.  
 Biol. Chem. Hoppe-Seyler 371, 423-431, 1990  
 A:Title: Identification of the aspartamide structure in a previously-reported peptide.  
 A:Reference number: S09685; MUID:90334745; PMID:2378679  
 A:Accession: S09685  
 A:Molecule type: protein  
 A:Residues: 27-34;17-36-47 <ROB>  
 R:De Vos, A.M.; Ullsch, M.; Kossiakoff, A.A.  
 Science 255, 306-312, 1992  
 A:Title: Human growth hormone and extracellular domain of its receptor: crystal structu  
 A:Reference number: A41728; MUID:92196577; PMID:1549776  
 A:Contents: annotation; X-ray crystallography, 2.8 angstroms  
 A:Note: The structure of the complex with growth hormone receptor is described  
 R:Gray, G.L.; Baldridge, J.S.; McKeown, K.S.; Heyneker, H.L.; Chang, C.N.  
 Gene 34, 247-254, 1985  
 A:Title: Periplasmic production of correctly processed human growth hormone in Escherich  
 A:Reference number: I41126; MUID:86137393; PMID:3912261  
 A:Accession: I84549  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-26 <RES>  
 A:Cross-references: GB:M14398; NID:q183158; PIDN:AAA52554.1; PID:q183159

C:Comment: The gene for this hormone is transcribed only in somatotrophic cells of t  
 C:Comment: About 90% of somatotropin is the 22K long form.  
 C:Genetics:  
 A:Gene: GDB:GH1  
 A:Cross-references: GDB:119982; OMIM:139250  
 A:Map position: 17q23.1-17q23.3  
 A:Introns: 4/1; 57/3; 97/3; 152/3  
 C:Superfamily: prolactin  
 C:Keywords: alternative splicing; hormone; pituitary  
 F:1-26/Domain: signal sequence #status predicted <SIG>  
 F:27-217/Product: somatotropin 1, long form #status experimental <SOL>  
 F:27-69/Product: growth hormone 5K peptide #status experimental <SKP>  
 F:27-73-217/Product: somatotropin 1, short form #status experimental <SOS>  
 F:79-151,208-215/Disulfide bonds: #status experimental  
 Query Match: 97.9% Score 460; DB 1; Length 217;  
 Best Local Similarity 46.94; Pred. No. 4.4e-42;  
 Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 2 FPTIPLSRLEFDNAMLRAHRLHQLAFDTYQFFREAYIPKEQKYSFLQNPOTLSLSESIPT 61  
 Db 27 FPTIPLSRLEFDNAMLRAHRLHQLAFDTYQFFREAYIPKEQKYSFLQNPOTLSLSESIPT 86  
 Qy 62 PSNRETOOKSNLELLIRISLLLIQSMLPEVQ 92  
 Db 87 PSNRETOOKSNLELLIRISLLLIQSMLPEVQ 117  
 RESULT 2  
 167410  
 Somatotropin - rhesus macaque  
 N:Alternate names: growth hormone  
 C:Species: Macaca mulatta (rhesus macaque)  
 C:Date: 31-May-1996 #sequence\_revision 11-May-1996 #text\_change 16-Jul-1999  
 C:Accession: 167410; A05094  
 R:Gollos, T.G.; Durning, M.; Fisher, J.M.; Fowler, P.D.  
 Endocrinology 133, 1744-1752, 1993  
 A:Title: Cloning of four growth hormone/chorionic somatomotropin-related compleme  
 A:Reference number: 153267; MUID:94008724; PMID:8404617  
 A:Accession: 167410  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-217 <RES>  
 A:Cross-references: GR:116556; NID:q293114; PIUN:AAA18842.1; PID:q293115  
 R:Li, C.H.; Chung, D.; Lahm, H.W.; Stein, S.  
 Arch. Biochem. Biophys. 245, 287-291, 1986  
 A:Title: The primary structure of monkey pituitary growth hormone.  
 A:Reference number: A05094; MUID:86129460; PMID:3080959  
 A:Accession: A05094  
 A:Molecule type: protein  
 A:Residues: 27-99;101-178;180-217 <LIC>  
 A:Note: the monkey species is not identified in the reference  
 R:Raben, M.S.  
 Science 125, 883-884, 1957  
 A:Title: Preparation of growth hormone from pituitaries of man and monkey.  
 A:Reference number: A4774  
 A:Contents: annotation; identification of source organism  
 C:Superfamily: prolactin  
 Query Match: 97.9% Score 460; DB 2; Length 217;  
 Best Local Similarity 46.94; Pred. No. 4.4e-42;  
 Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 2 FPTIPLSRLEFDNAMLRAHRLHQLAFDTYQFFREAYIPKEQKYSFLQNPOTLSLSESIPT 61  
 Db 27 FPTIPLSRLEFDNAMLRAHRLHQLAFDTYQFFREAYIPKEQKYSFLQNPOTLSLSESIPT 86  
 Qy 62 PSNRETOOKSNLELLIRISLLLIQSMLPEVQ 92  
 Db 87 PSNRETOOKSNLELLIRISLLLIQSMLPEVQ 117  
 RESULT 3

STHUV  
somatotropin 2 precursor - human  
N:Alternate names: growth hormone 2; growth hormone variant; hGH-V; placental somatotropin  
N:Contains: somatotropin 2, long splice form; somatotropin 2, short splice form  
C:Species: homo sapiens (man)  
C:Date: 17-Dec-1982 #sequence\_revision 10-Feb-1995 #text\_change 21-Jul-2000  
C:Accession: D12435; B28072; A152104; A60711  
R:Chen, E.Y.; Liao, Y.C.; Smith, D.H.; Barrera-Saldana, H.A.; Gelinas, K.E.; Seeburg, P.  
Genomics 4: 479-497, 1989  
A:Title: The human growth hormone locus: nucleotide sequence, biology, and evolution.  
A:Reference number: A32435; MUID:89307277; PMID:2744760  
A:Accession: D12435  
A:Molecule type: DNA  
A:Residues: 1-217 <CHE>  
A:Cross-references: GB:J03071; NID:q183148; PIDN:AAA52552.1; PID:q183152  
R:Cooke, N.E.; Kay, J.; Emery, J.G.; Liebhauer, S.A.  
J. Biol. Chem. 263: 9001-9006, 1988  
A:Title: Two distinct species of human growth hormone variant mRNA in the human placenta  
A:Reference number: A92725; MUID:88243759; PMID:3379057  
A:Accession: B28072  
A:Molecule type: mRNA  
A:Residues: 1-217 <COO>  
A:Note: an alternative splice junction for intron 4 is used  
C:Genetics:  
A:Gene: GH2; GH2  
A:Cross-references: GDB:119943; OMIM:139240  
A:Map position: 17q22-17q24  
A:Introns: 4/1; 57/3; 97/3; 152/3  
C:Superfamily: prolactin  
C:Keywords: alternative splicing; hormone; placenta  
F:1-26/Domain: signal sequence #status predicted <SIG>  
F:27-217/Product: somatotropin 2 splice form 2 #status predicted <MAI>  
Query Match 89.8%; Score 422; DB 1; Length 256;  
Best Local Similarity 92.3%; Pred. No. 6.7e-38;  
Matches 84; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
A:Accession: A01511  
A:Molecule type: DNA  
A:Residues: 1-34; p. 36-217 <SEE>  
R:Igout, A.; Scippo, M.L.; Frankenne, F.; Henneen, G.  
Arch. Int. Physiol. Biochim. 96: 63-67, 1988  
A:Title: Cloning and nucleotide sequence of placental hGH-V cDNA.  
A:Reference number: 152104; MUID:89024984; PMID:2460050  
A:Accession: 152104  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-217 <ICO>  
A:Cross-references: GB:M38451; NID:q183179; L: N:AAA35891.1; PID:q183180  
R:Frankenne, F.; Scippo, M.L.; Van Beeumen, J.; Igout, A.; Henneen, G.  
J. Clin. Endocrinol. Metab. 71: 15-18, 1990  
A:Title: Identification of placental human growth hormone as the growth hormone-V gene  
A:Reference number: A60711; MUID:90317018; PMID:2196278  
A:Accession: A60711  
A:Molecule type: protein  
A:Residues: 27-44; 46-57 <FRA>  
A:Experimental source: tissue placenta  
A:Note: partial glycosylation was demonstrated by lectin binding  
C:Comment: This gene is expressed by the placenta.  
C:Genetics:  
A:Gene: GH2; GH2  
A:Cross-references: GDB:119943; OMIM:139240  
A:Map position: 17q22-17q24  
A:Introns: 4/1; 57/3; 97/3; 152/3  
C:Superfamily: prolactin  
C:Keywords: alternative splicing; glycoprotein; hormone; placenta  
F:1-26/Domain: signal sequence #status predicted <SIG>  
F:27-217/Product: somatotropin 2, long splice form #status predicted <SIG>  
F:27-57.73-217/Product: somatotropin 2, short splice form #status predicted <SIG>  
F:79-191-208-215/Disulfide bonds: #status predicted  
F:166/Binding site: carbohydrate (Asn) (covalent) #status predicted  
Query Match 89.8%; Score 422; DB 1; Length 217;  
Best Local Similarity 92.3%; Pred. No. 5.9e-38;  
Matches 84; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 2 FPTPLSLFDFNAMLRAHRLHQLAFTDYQEFEEAYIPKEOKYSLFONPOTSLSFSES IPT 61  
DB 27 FPTPLSLFDFNAMLRAHRLHQLAFTDYQEFEEA LKEOKYSLFONPOTSLSFSES IPT 86  
QY 62 PSNREETOOKSNLELLRISLLLIQSLEVPVQ 92  
DB 87 PSNRVKTOKSNLELLRISLLLIQSLEVPVQ 117

## RESULT 4

STHUV2  
somatotropin 2 precursor, splice form 2 - human  
N:Alternate names: growth hormone variant-2; placental somatotropin form 2  
C:Species: homo sapiens (man)  
C:Date: 30-Sep-1989 #sequence\_revision 10-Feb-1995 #text\_change 02-Sep-1997  
C:Accession: A28072  
R:Cooke, N.E.; Kay, J.; Emery, J.G.; Liebhauer, S.A.  
J. Biol. Chem. 263: 9001-9006, 1988  
A:Title: Two distinct species of human growth hormone variant mRNA in the human placenta  
A:Reference number: A92725; MUID:88243759; PMID:3379057  
A:Accession: A28072  
A:Molecule type: mRNA  
A:Residues: 1-256 <COO>  
A:Note: an alternative splice junction for intron 4 is used  
C:Genetics:  
A:Gene: GH2; GH2  
A:Cross-references: GDB:119943; OMIM:139240  
A:Map position: 17q22-17q24  
A:Introns: 4/1; 57/3; 97/3; 152/3  
C:Superfamily: prolactin  
C:Keywords: alternative splicing; hormone; placenta  
F:1-26/Domain: signal sequence #status predicted <SIG>  
F:27-256/Product: somatotropin 2 splice form 2 #status predicted <MAI>  
Query Match 89.8%; Score 422; DB 1; Length 256;  
Best Local Similarity 92.3%; Pred. No. 6.7e-38;  
Matches 84; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 2 FPTPLSLFDFNAMLRAHRLHQLAFTDYQEFEEAYIPKEOKYSLFONPOTSLSFSES IPT 61  
DB 27 FPTPLSLFDFNAMLRAHRLHQLAFTDYQEFEEAYIPKEOKYSLFONPOTSLSFSES IPT 86  
QY 62 PSNREETOOKSNLELLRISLLLIQSLEVPVQ 92  
DB 87 PSNRVKTOKSNLELLRISLLLIQSLEVPVQ 117  
RESULT 5  
167411  
somatotropin - rhesus macaque  
N:Alternate names: growth hormone  
C:Species: Macaca mulatta (rhesus macaque)  
C:Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 16-Jul-1999  
C:Accession: 167411  
R:Colos, J.G.; Durning, M.; Fisher, J.M.; Fowler, P.D.  
Endocrinology 133: 1744-1752, 1993  
A:Title: Cloning of four growth hormone/chorionic somatomammotropin related comple  
A:Reference number: 153267; MUID:94008724; PMID:8404617  
A:Accession: 167411  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-217 <RES>  
A:Cross-references: GB:116555; NID:q293116; PIDN:AAA20180.1; PID:q293117  
C:Superfamily: prolactin  
Query Match 85.5%; Score 402; DB 2; Length 217;  
Best Local Similarity 85.7%; Pred. No. 7.8e-46;  
Matches 78; Conservative 6; Mismatches 7; Indels 0; Gaps 0;  
QY 2 FPTPLSLFDFNAMLRAHRLHQLAFTDYQEFEEAYIPKEOKYSLFONPOTSLSFSES IPT 61  
DB 27 FPTPLSLFDFNAMLRAHRLHQLAFTDYQEFEEAYIPKEOKYSLFONPOTSLSFSES IPT 86  
QY 62 PSNREETOOKSNLELLRISLLLIQSLEVPVQ 92  
DB 87 PSNREETOOKSNLELLRISLLLIQSLEVPVQ 117  
RESULT 6  
167409  
chorionic somatomammotropin-3 - rhesus macaque

C:Species: Macaca mulatta (rhesus macaque)  
 C:Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 16-Jul-1999  
 C:Accession: 167409  
 R:Colos, T.G.; Durning, M.; Fisher, J.M.; Fowler, P.D.  
 Endocrinology 133, 1744-1752, 1993  
 A:Title: Cloning of four growth hormone/chorionic somatomotropin-related complementary DNAs from the rhesus macaque  
 A:Reference number: 153267; MUID:94008724; PMID:8404617  
 C:Accession: 167409  
 A:Status: preliminary; translated from GR/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-217 <RES>  
 A:Cross-references: GB:L16554; NID:q293112; PIDN:AAA18841.1; PID:q293113  
 C:Superfamily: prolactin

Query Match 84.1%; Score 396; DB 2; Length 217;  
 Best Local Similarity 82.2%; Pred. No. 3.4e-35;  
 Matches 74; Conservative 11; Mismatches 5; Indels 0; Gaps 0;  
 QY 3 PTPLSLRFDNMLRAHRIHQALAFDTYQFEAYIPKEOKYFSLONPOTSLSFSESITPT 62  
 DB 28 PSVPLSKLFDMIAHQHRIHQALAFDTYQFEAYIPKEKKHSLMGNQASFCFADSIPT 87  
 QY 63 SNRETOQKSNLELLRLISLLIQSWLEPVQ 92  
 DB 88 SNRETOQKSNLELLRLISLLIQSWLEPVQ 112

RESULT 7  
 167408  
 chorionic somatomotropin-2 - rhesus macaque (fragment)  
 C:Species: Macaca mulatta (rhesus macaque)  
 C:Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 16-Jul-1999  
 C:Accession: 167408  
 R:Colos, T.G.; Durning, M.; Fisher, J.M.; Fowler, P.D.  
 Endocrinology 133, 1744-1752, 1993  
 A:Title: Cloning of four growth hormone/chorionic somatomotropin-related complementary DNAs from the rhesus macaque  
 A:Reference number: 153267; MUID:94008724; PMID:8404617  
 C:Accession: 167408  
 A:Status: preliminary; translated from GR/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-212 <RES>  
 A:Cross-references: GB:L16553; NID:q293110; PIDN:AAA18840.1; PID:q293111  
 C:Superfamily: prolactin

Query Match 84.1%; Score 396; DB 2; Length 212;  
 Best Local Similarity 82.2%; Pred. No. 3.4e-35;  
 Matches 74; Conservative 11; Mismatches 5; Indels 0; Gaps 0;  
 QY 3 PTPLSLRFDNMLRAHRIHQALAFDTYQFEAYIPKEOKYFSLONPOTSLSFSESITPT 62  
 DB 23 PSVPLSKLFDMIAHQHRIHQALAFDTYQFEAYIPKEKKHSLMGNQASFCFADSIPT 82  
 QY 63 SNRETOQKSNLELLRLISLLIQSWLEPVQ 92  
 DB 83 SNRETOQKSNLELLRLISLLIQSWLEPVQ 112

RESULT 6  
 153267  
 chorionic somatomotropin-1 - rhesus macaque  
 C:Species: Macaca mulatta (rhesus macaque)  
 C:Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 16-Jul-1999  
 C:Accession: 153267  
 R:Colos, T.G.; Durning, M.; Fisher, J.M.; Fowler, P.D.  
 Endocrinology 133, 1744-1752, 1993  
 A:Title: Cloning of four growth hormone/chorionic somatomotropin-related complementary DNAs from the rhesus macaque  
 A:Reference number: 153267; MUID:94008724; PMID:8404617  
 C:Accession: 153267  
 A:Status: preliminary; translated from GR/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-217 <RES>  
 A:Cross-references: GB:L16552; NID:q293108; PIDN:AAA18839.1; PID:q293109  
 C:Superfamily: prolactin

Query Match 84.1%; Score 396; DB 2; Length 217;  
 Best Local Similarity 82.2%; Pred. No. 3.4e-35;  
 Matches 74; Conservative 11; Mismatches 5; Indels 0; Gaps 0;  
 QY 3 PTPLSLRFDNMLRAHRIHQALAFDTYQFEAYIPKEOKYFSLONPOTSLSFSESITPT 62  
 DB 28 PSVPLSKLFDMIAHQHRIHQALAFDTYQFEAYIPKEKKHSLMGNQASFCFADSIPT 87  
 QY 63 SNRETOQKSNLELLRLISLLIQSWLEPVQ 92  
 DB 88 SNRETOQKSNLELLRLISLLIQSWLEPVQ 117

RESULT 9  
 153192  
 chorionic somatomotropin A precursor (validated) - human  
 N:Alternative names: chorionic somatomotropin; placental lactogen  
 C:Species: Homo sapiens (man)  
 C:Date: 23-Oct-1981 #sequence\_revision 23-Oct-1981 #text\_change 08-Dec-2000  
 C:Accession: C32435; A94422; 152342; A93833; A93192; A90054; A94427; A61283; 155229;  
 R:Chen, E.Y.-J.; Liaw, Y.-C.; Smith, D.H.; Barrera-Saldana, H.A.; Gellinas, R.E.; Seeburg,  
 Genomics 4, 479-497, 1989  
 A:Title: The human growth hormone locus: nucleotide sequence, biology, and evolution  
 A:Reference number: A32435; MUID:89407277; PMID:2744760  
 C:Accession: C32435  
 A:Molecule type: DNA  
 A:Residues: 1-217 <CHE>  
 A:Cross-references: GB:J04371; NID:q183148; PIDN:AAA52551.1; PID:q183151  
 R:Goodman, H.M.; DeNoto, F.; Fiddes, J.C.; Hillebrand, R.A.; Papk, G.S.; Smith, S.; T.  
 in Mobilization and Reassembly of Genetic Information, Scott, W.A.; Werner, R.; Josep  
 A:Reference number: A94422  
 C:Accession: A94422  
 A:Molecule type: mRNA  
 A:Residues: 1-217 <GCO>  
 R:Tanaka, M.; Masuda, N.; Watahiki, M.; Yamakawa, M.; Shimizu, K.; Nagai, J.; Nakash;  
 Biochem. Int. 16, 287-292, 1988  
 A:Title: cDNA cloning of human chorionic somatomotropin-1 mRNA whose transcription  
 A:Reference number: 152342; MUID:88209096; PMID:2835050  
 C:Accession: 152342  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-3 <TAN>  
 R:Sherwood, L.M.; Bursstein, Y.; Schechter, I.  
 Proc. Natl. Acad. Sci. U.S.A. 76, 3819-3823, 1979  
 A:Title: Primary structure of the NH-2-terminal extra piece of the precursor to human  
 A:Reference number: A93833; MUID:80034970; PMID:291043  
 C:Accession: A93833  
 A:Molecule type: protein  
 A:Residues: 1-3-26 <SHE>  
 A:Experimental source: placenta  
 R:Shine, J.; Seeburg, P.H.; Martial, J.A.; Baxter, J.D.; Goodman, H.M.  
 Nature 270, 494-499, 1977  
 A:Title: Construction and analysis of recombinant DNA for human chorionic somatomamm  
 A:Reference number: A93192; MUID:78071761; PMID:593368  
 C:Accession: A93192  
 A:Molecule type: DNA  
 A:Residues: 50-217 <SHI>  
 R:Li, C.H.; Dixon, J.S.; Chung, D.  
 Arch. Biochem. Biophys. 155, 95-110, 1973  
 A:Title: Amino acid sequence of human chorionic somatomammotropin.  
 A:Reference number: A90054; MUID:73201971; PMID:4712450  
 C:Accession: A90054  
 A:Molecule type: protein  
 A:Residues: 27-217 <LIC>  
 A:Experimental source: placenta  
 R:Niall, H.D.  
 in Prolactin and Carcinogenesis, Proc. Fourth Tenovus Workshop Prolactin, Griffiths,  
 A:Title: The chemistry of the human lactogenic hormones.  
 A:Reference number: A94427  
 C:Accession: A94427

RESULT 12  
B49159  
somatotropin - golden hamster  
N:Alternate names: growth hormone  
C:Species: *Mesocricetus auratus* (golden hamster)  
C:Date: 19-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 21-Jul-2000  
C:Accession: B49159  
R:Southard, J.N.; Sanchez-Jimenez, F.; Campbell, G.T.; Talamantes, F.  
Endocrinology 129, 2965-2971, 1991  
A:Title: Sequence and expression of hamster prolactin and growth hormone messenger  
A:Reference number: A49159; MUID:92063850; PMID:1954881



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:54:00; Search time 8.57348 seconds  
(without alignments)  
504.633 Million cell updates/sec

Title: US-09-423-100-2

Perfect score: 470

Sequence: 1 MFPTIPLSRLEFDNAMI RARR.....NLELRISILLTOSWLEFVU 92

Scoring table: BLAST2  
Gap: 10.0, Gap ext: 0.7

Searched: 127865 seqs, 47,463,065 total hits

Total number of hits satisfying chosen parameters: 127865

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Swissprot\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result NO.	Score	Query Match	Length	DB ID	Description
1	46.0	97.9	217	1	SOMA_HUMAN
2	46.0	97.9	217	1	SOMA_MACHU
3	46.0	97.9	217	1	SOMA_PANTR
4	44.4	92.1	217	1	SOMA_SALIB
5	43.2	91.7	217	1	SOMA_CALJA
6	43.0	91.5	217	1	SOMA_PANTR
7	42.2	89.8	217	1	SOMA_HUMAN
8	42.2	89.8	217	1	SOMA_HUMAN
9	41.1	81.1	217	1	SOMA_HUMAN
10	41.1	81.1	217	1	SOMA_HUMAN
11	40.5	66.1	216	1	SOMA_MESAJ
12	40.5	65.4	190	1	SOMA_BALRC
13	40.5	64.8	216	1	SOMA_MOUSE
14	40.5	64.4	216	1	SOMA_MOUSE
15	40.5	64.4	216	1	SOMA_HABIT
16	40.5	64.4	216	1	SOMA_RAT
17	40.5	64.4	217	1	SOMA_MOUSE
18	40.5	64.1	190	1	SOMA_MOUSE
19	40.5	64.1	216	1	SOMA_MOUSE
20	40.5	64.1	216	1	SOMA_MOUSE
21	40.5	64.1	216	1	SOMA_MOUSE
22	40.5	64.1	216	1	SOMA_MOUSE
23	40.5	64.1	216	1	SOMA_MOUSE
24	40.5	64.1	216	1	SOMA_MOUSE
25	40.5	64.1	216	1	SOMA_MOUSE
26	40.5	64.1	216	1	SOMA_MOUSE
27	40.5	64.1	216	1	SOMA_MOUSE
28	40.5	64.1	216	1	SOMA_MOUSE
29	40.5	64.1	216	1	SOMA_MOUSE
30	40.5	64.1	216	1	SOMA_MOUSE
31	40.5	64.1	216	1	SOMA_MOUSE
32	40.5	64.1	216	1	SOMA_MOUSE
33	40.5	64.1	216	1	SOMA_MOUSE

34 272.5 58.0 190 1 SOMA\_CRONO  
35 268.5 57.1 191 1 SOMA\_CHEMY  
36 261 55.5 216 1 SOMA\_ANAPL  
37 257.5 54.8 190 1 SOMA\_ACIGU  
38 257.5 54.8 190 1 SOMA\_ACIGU  
39 247.5 52.7 211 1 SOMA\_LEPOS  
40 239.5 51.0 214 1 SOMA\_XENLA  
41 238.5 50.7 215 1 SOMA\_RANCA  
42 226.5 48.2 213 1 SOMA\_BUFMA  
43 225.5 48.0 183 1 SOMA\_PRIGL  
44 219.5 46.7 208 1 SOMA\_XENLA  
45 218.5 46.5 206 1 SOMA\_PROAN

#### ALIGNMENTS

RESULT 1  
SOMA\_HUMAN SEABARDI FRT 217 AA  
AC P01241 Q24405 Q16641 Q98671 Q50M37 Q60M55  
DT 21-JUL-1986 (Ref: 01, Created)  
DT 01-MAR-1992 (Ref: 21, Last sequence update)  
DE 15-SEP-2003 (Ref: 47, Last annotation update)  
DE Somatotropin precursor (Growth hormone) (GH) (GH-N) (Pituitary growth hormone) (Growth hormone 1).  
GN GH1  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID:9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE:80034477; PubMed:386281;  
RA Roskam W., Roudeon F.;  
RT "Molecular cloning and nucleotide sequence of the human growth hormone structural gene";  
RL Nucleic Acids Res. 7:305-320(1979).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE:79201293; PubMed:377496;  
RA Martini J.A., Hallowell R.A., Baxter J.D., Goodman H.M.;  
RT "Human growth hormone: complementary DNA cloning and expression in bacteria";  
RL Science 205:602-607(1979).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 1), AND POSSIBLE ALTERNATIVE SPLICING.  
RX MEDLINE:82014939; PubMed:6269091;  
RA Denoto F.M., Moore D.D., Goodman H.M.;  
RT "Human growth hormone DNA sequence and mRNA structure: possible alternative splicing";  
RL Nucleic Acids Res. 9:3719-3730(1981).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE:83182010; PubMed:7169009;  
RA Seeburg P.H.;  
RT "The human growth hormone gene family: nucleotide sequences show recent divergence and predict a new polypeptide hormone";  
RL DNA 1:239-249(1982).  
RN [5]  
RP SEQUENCE FROM N.A.  
RX MEDLINE:89307277; PubMed:2744760;  
RA Chen E.Y., Liao Y.C., Smith D.H., Barrera-Saldana H.A.,  
RA Gellinas R.E., Seeburg P.H.;  
RT "The human growth hormone locus: nucleotide sequence, biology, and evolution";  
RL Genomics 4:479-497(1989).  
RN [6]  
RP SEQUENCE FROM N.A. (ISOFORM 3).  
RC TISSUE-Pituitary;  
RA Gu J., Huang Q.-H., Li N., Xu S.-H., Han Z. G., Fu G., Chen Z.;  
RT "A novel gene expressed in human pituitary";  
RN Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases



Query Match 97.9%; Score 460; DR 1; Length 217;  
 Best Local Similarity 98.9%; Pred. No. 8.8e-41;  
 Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 FPTIPLSRFDNAMLRAHRLHQLAFDTYOEFEEAYIPKEQKYSFLQNPOTSLCFSES IPT 61  
 Db 27 FPTIPLSRFDNAMLRAHRLHQLAFDTYOEFEEAYIPKEQKYSFLQNPOTSLCFSES IPT 86  
 OY 62 PSNREETOQSNLELLRLISLLLSQSWLEPVQ 92  
 Db 87 PSNREETOQSNLELLRLISLLLSQSWLEPVQ 117

RESULT 2

SOMA\_MACMO STANDARD: PRT: 217 AA

AC P33094;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1994 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Somatotropin precursor (Growth hormone) (GH) (GH-N) (Pituitary growth hormone) (Growth hormone 1).  
 GN GH1.  
 OS Macaca mulatta (Rhesus macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Reptbrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecoidea; Macaca.  
 OX NCBI\_TaxID=9544;  
 {}  
 {}  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=34008724; PubMed=8404617;  
 RA Golos T.G., Durning M., Fisher J.M., Fowler P.D.;  
 RT "Cloning of four growth hormone/chorionic somatomotropin-related complementary deoxyribonucleic acids differentially expressed during pregnancy in the rhesus monkey placenta.";  
 RL Endocrinology 133:1744-1752(1993).  
 RN [2]  
 RP SEQUENCE OF 27-217.  
 RX MEDLINE=86129460; PubMed=3080959;  
 RA Li C.H., Chung D., Lahn H.W., Stein S.;  
 RT "The primary structure of monkey pituitary growth hormone.";  
 RL Arch. Biochem. Biophys. 245:287-291(1986).  
 CC -1- FUNCTION: Plays an important role in growth control. Its major role in stimulating body growth is to stimulate the liver and other tissues to secrete IGF-1. It stimulates both the differentiation and proliferation of myoblasts. It also stimulates amino acid uptake and protein synthesis in muscle and other tissues.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: L16556; AAA18842.1;  
 DR F1R: 167410; 167410.  
 DR HSSP: P01241; 1AXI.  
 DR InterPro: IPR001400; Somatotropin.  
 DR Pfam: PF00103; hormone.1  
 DR PRINTS: PR00836; SOMATOTROPIN.  
 DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
 DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
 KW Hormone; Pituitary; Signal.  
 FT SIGNAL 1 26  
 FT CHAIN 27 217 SOMATOTROPIN.  
 FT DISULFID 79 191 BY SIMILARITY.  
 FT DISULFID 208 215 BY SIMILARITY.  
 FT CONFLICT 100 100 SOMATOTROPIN.  
 FT DISULFID 79 191 BY SIMILARITY.  
 FT DISULFID 208 215 BY SIMILARITY.  
 FT CONFLICT 100 100 E -> Q (IN REF. 2).

FT CONFLICT 179 179 N -> D (IN REF. 2).  
 SO SEQUENCE 217 AA: 24913 MW: 2C5180341EFC46D0 CRC64;

Query Match 97.9%; Score 460; DB 1; Length 217;  
 Best Local Similarity 98.9%; Pred. No. 8.8e-41;  
 Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 FPTIPLSRFDNAMLRAHRLHQLAFDTYOEFEEAYIPKEQKYSFLQNPOTSLCFSES IPT 61  
 Db 27 FPTIPLSRFDNAMLRAHRLHQLAFDTYOEFEEAYIPKEQKYSFLQNPOTSLCFSES IPT 86  
 OY 62 PSNREETOQSNLELLRLISLLLSQSWLEPVQ 92  
 Db 87 PSNREETOQSNLELLRLISLLLSQSWLEPVQ 117

RESULT 3

SOMA\_PANIP STANDARD: PRT: 217 AA

AC P38746;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Somatotropin precursor (Growth hormone) (GH) (GH-N) (Pituitary growth hormone) (Growth hormone 1).  
 GN GH1.  
 OS Pan troglodytes (Chimpanzee).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
 OX NCBI\_TaxID=9598;  
 {}  
 {}  
 RN SEQUENCE FROM N.A.  
 RP Revol A., Esquivel D., Santiago D., Barrera Saldana H.;  
 RA "Independent duplication of the growth hormone gene in three Anthropoid lineages.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Plays an important role in growth control. Its major role in stimulating body growth is to stimulate the liver and other tissues to secrete IGF-1. It stimulates both the differentiation and proliferation of myoblasts. It also stimulates amino acid uptake and protein synthesis in muscle and other tissues (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: AF374232; AAL72284.1;  
 DR InterPro: IPR001400; Somatotropin.  
 DR Pfam: PF00103; hormone.1.  
 DR PRINTS: PR00836; SOMATOTROPIN.  
 DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
 DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
 KW Hormone; Pituitary; Signal.  
 FT SIGNAL 1 26 BY SIMILARITY.  
 FT CHAIN 27 217 SOMATOTROPIN.  
 FT DISULFID 79 191 BY SIMILARITY.  
 FT DISULFID 208 215 BY SIMILARITY.  
 SO SEQUENCE 217 AA: 24843 MW: FEA295EDEF0518674 CRC64;

Query Match 97.9%; Score 460; DB 1; Length 217;  
 Best Local Similarity 98.9%; Pred. No. 8.8e-41;  
 Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 FPTIPLSRFDNAMLRAHRLHQLAFDTYOEFEEAYIPKEQKYSFLQNPOTSLCFSES IPT 61  
 Db 27 FPTIPLSRFDNAMLRAHRLHQLAFDTYOEFEEAYIPKEQKYSFLQNPOTSLCFSES IPT 86

```

QY 62 PSNRETOOKSNLELLRISLLIQSWLEPVQ 2
DB 87 PSNRETOOKSNLELLRISLLIQSWLEPVQ 117

RESULT 4
SOMA_SA:SB
ID SOMA_SA:SB STANDARD; PRT: 217 AA.
AC P58343;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Somatotropin precursor (Growth hormone).
GN Callithrix jacchus (Common marmoset).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;
OX NCBI_TaxID=9483;
RN [1];
RP SEQUENCE FROM N.A.
RA Liu J.C., Makova K.D., Adkins R.M., Gibson S., Li W.H.;
RT "Epistatic evolution of growth hormone in primates and emergence of the
RL species specificity of human growth hormone receptor.";
CC Mol. Biol. Evol. 18:945-953(2001)
CC -1- FUNCTION: Plays an important role in growth control. Its major
CC role in stimulating body growth is to stimulate the liver and
CC other tissues to secrete IGF-1. It stimulates both the
CC differentiation and proliferation of myoblasts. It also stimulates
CC amino acid uptake and protein synthesis in muscle and other
CC tissues (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF349060; AAK62287.1;
CC InterPro: IPR001400; Somatotropin.
CC Pfam: PF00103; hormone; 1.
CC PRINTS: PS00836; SOMATOTROPIN.
CC PROSITE: PS00266; SOMATOTROPIN_1; 1.
CC PROSITE: PS00338; SOMATOTROPIN_2; 1.
CC Hormone; Pituitary; Signal.
CC SIGNAL: 1 26 BY SIMILARITY.
CC CHAIN: 27 217 SOMATOTROPIN.
CC DISULFID: 79 191 BY SIMILARITY.
CC DISULFID: 208 215 BY SIMILARITY.
CC SEQUENCE: 217 AA; 24864 MW; 951528999.52997 CRC64.

Query Match 92.1%; Score 433; DB 1; Length 217;
Best Local Similarity 91.2%; Pred. No. 5,6e-38;
Matches 81; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 FPTPLSRLEFDNAMIKAHRLHQAFDTYQFFEAYIPKEQKYSFLONPQTSLSFSESPT 61
DB 27 FPTPLSRLEFDNAMIKAHRLHQAFDTYQFFEAYIPKEQKYSFLONPQTSLSFSESPT 86

QY 62 PSNRETOOKSNLELLRISLLIQSWLEPVQ
DB 87 PSNRETOOKSNLELLRISLLIQSWLEPVQ .7

RESULT 5
SOMA_CALJA
ID SOMA_CALJA STANDARD; PRT: 217 AA.
AC Q9GHB3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Growth hormone variant precursor (GH-V) (Placenta-specific growth
DE hormone) (Growth hormone 2).
GN GH2.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9598;

```

```

DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Somatotropin precursor (Growth hormone).
GN Callithrix jacchus (Common marmoset).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;
OX NCBI_TaxID=9483;
RN [1];
RP SEQUENCE FROM N.A.
RA Wallis G.C., Wallis M.;
RT "Cloning and characterisation of a putative growth hormone encoding
RT gene from the marmoset (Callithrix jacchus).";
CC Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Plays an important role in growth control. Its major
CC role in stimulating body growth is to stimulate the liver and
CC other tissues to secrete IGF-1. It stimulates both the
CC differentiation and proliferation of myoblasts. It also stimulates
CC amino acid uptake and protein synthesis in muscle and other
CC tissues (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AJ297563; CAC03431.1;
CC HSP: P01241; 1A22.
CC InterPro: IPR001400; Somatotropin.
CC Pfam: PF00103; hormone; 1.
CC PROSITE: PS00266; SOMATOTROPIN_1; 1.
CC PROSITE: PS00338; SOMATOTROPIN_2; 1.
CC Hormone; Pituitary; Signal.
CC SIGNAL: 1 26 BY SIMILARITY.
CC CHAIN: 27 217 SOMATOTROPIN.
CC DISULFID: 79 191 BY SIMILARITY.
CC DISULFID: 208 215 BY SIMILARITY.
CC SEQUENCE: 217 AA; 24959 MW; 8102151A12CE6192 CRC64.

Query Match 91.9%; Score 432; DB 1; Length 217;
Best Local Similarity 91.2%; Pred. No. 7.2e-48;
Matches 83; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 FPTPLSRLEFDNAMIKAHRLHQAFDTYQFFEAYIPKEQKYSFLONPQTSLSFSESPT 61
DB 27 FPTPLSRLEFDNAMIKAHRLHQAFDTYQFFEAYIPKEQKYSFLONPQTSLSFSESPT 86

QY 62 PSNRETOOKSNLELLRISLLIQSWLEPVQ 92
DB 87 PSNRETOOKSNLELLRISLLIQSWLEPVQ 117

RESULT 6
SOM2_PANTR
ID SOM2_PANTR STANDARD; PRT: 217 AA.
AC P58757;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Growth hormone variant precursor (GH-V) (Placenta-specific growth
DE hormone) (Growth hormone 2).
GN GH2.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9598;

```





RT \*Analysis of a major human chorionic somatomammotropin gene. Evidence  
 RT for two functional promoter elements.\*;  
 RL J. Biol. Chem. 259:13131-13138(1984).  
 RN [2]  
 RP SEQUENCE FROM N.A. (GENE CS#2)  
 RX MEDLINE=87161235; PubMed=3030680;  
 RA Hirt H., Krimelman J., Birnbaum M.J., Chen E.Y., Seeburg P.H.,  
 RA Eberhardt N.L., Barta A.;  
 RT \*The human growth hormone gene locus: structure, evolution, and  
 RT allelic variations.\*;  
 RL DNA 6:59-70(1987).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=83160916; PubMed=6300056;  
 RA Barrera-Saldana H.A., Seeburg P.H., Saunders G.F.;  
 RT \*Two structurally different genes produce the same secreted human  
 RT placental lactogen hormone.\*;  
 RL J. Biol. Chem. 258:3787-3794(1983).  
 RN [4]  
 RP SEQUENCE FROM N.A. (GENES CS#1 AND CS#2).  
 RX MEDLINE=89407277; PubMed=2744760;  
 RA Chen E.Y., Liao Y.C., Smith D.H., Barrera-Saldana H.A., Golligorsky E.,  
 RA Seeburg P.H.;  
 RT \*The human growth hormone locus: nucleotide sequence, biology, and  
 RT evolution.\*;  
 RL Genomics 4:479-497(1989).  
 RN [5]  
 RP SEQUENCE.  
 RX MEDLINE=83482010; PubMed 7169009;  
 RA Seeburg P.H.;  
 RT \*The human growth hormone gene family: nucleotide sequences show  
 RT recent divergence and predict a new polypeptide hormone.\*;  
 RL DNA 1:239-249(1982).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta, and Uterus;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,  
 RA Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong E.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Soderstrom E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smolius D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT \*Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.\*;  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [7]  
 RP SEQUENCE OF 50-217 FROM N.A.  
 RX MEDLINE=78071761; PubMed=593368;  
 RA Shue J., Seeburg P.H., Martial J.A., Baxter J.D., Goodman H.M.;  
 RT \*Construction and analysis of recombinant DNA for human chorionic  
 RT somatomammotropin.\*;  
 RL Nature 270:494-499(1977).  
 RN [8]  
 RP SEQUENCE OF 27-217  
 RX MEDLINE=73201971; PubMed=4712450;  
 RA Li C.H., Dixon J.S., Chung D.;  
 RT \*Amino acid sequence of human chorionic somatomammotropin.\*;  
 RL Arch. Biochem. Biophys. 155:95-110(1973).  
 RN [9]  
 RP SEQUENCE OF 27-117  
 RX MEDLINE=72016313; PubMed=5286363;

RA Sherwood L.M., Handwerger S., McLaurin W.D., Lanner M.;  
 RT \*Amino acid sequence of human placental lactogen.\*;  
 RL Nature New Biol. 233:59-61(1971).  
 RN [10]  
 RP ERRATUM.  
 RA Sherwood L.M., Handwerger S., McLaurin W.D., Lanner M.;  
 RL Nature New Biol. 235:64-64(1972).  
 RN [11]  
 RP INTERCHAIN DISULFIDE BONDS  
 RX MEDLINE=79173081; PubMed=438159;  
 RA Schneider A.B., Kowalski K., Russell J., Sherwood L.M.;  
 RT \*Identification of the interchain disulfide bonds of dimeric human  
 RT placental lactogen.\*;  
 RL J. Biol. Chem. 254:3782-3787(1979).  
 CC - FUNCTION: SIMILAR TO THAT OF SOMATOTROPIN.  
 CC - SUBCELLULAR LOCATION: Secreted.  
 CC - MISCELLANEOUS: THE SEQUENCE OF CS#1 IS SHOWN.  
 CC - SIMILARITY: BELONGS TO THE SOMATOTROPIN/PLACENTALIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: J00573; CAA23836.1;  
 DR EMBL: J00289; AAA58747.1;  
 DR EMBL: K02401; AAA52115.1;  
 DR EMBL: M15894; AAA52116.1;  
 DR EMBL: J03071; AAA52551.1;  
 DR EMBL: J00118; AAA98621.1;  
 DR EMBL: BC027177; AAH02717.1;  
 DR EMBL: BC005921; AAH05921.1;  
 DR EMBL: BC020756; AAH20756.1;  
 DR PIR: A26449; A26449.  
 DR PIR: C12435; LCHUC.  
 DR HSP: P01241; I422.  
 DR Genew; HGNC:2440; CS#1.  
 DR Genew; HGNC:2441; CS#2.  
 DR MIM: 150200;  
 DR GO: GO:0007565; P:pregnancy; TAS.  
 DR InterPro: IPR001400; Somatotropin.  
 DR Pfam: PF0104; hormone\_1  
 DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
 DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
 DR Hormone; Placenta; Multigene family; Signal.  
 FT SIGNAL 1 25  
 FT CHAIN 27 217 LACTOGEN.  
 FT DISULFID 79 131  
 FT DISULFID 208 215  
 FT DISULFID 208 208  
 FT DISULFID 215 215  
 FT VARIANT 3 3  
 FT VARIANT 104 105  
 FT CONFLICT 84 84  
 FT CONFLICT 95 95  
 FT CONFLICT 116 116  
 FT CONFLICT 134 136  
 FT CONFLICT 217 217  
 FT SEQUENCE 217 AA: 25020 MW: 23580DC7A711F431 CRC64;  
 Query Match 41.1%; Score 381; DB 1; Length 217;  
 Best Local Similarity 82.0%; Pred. No. 1.4e-32;  
 Matches 73; Conservative 8; Mismatches 8; Indels 0; Gaps 0;  
 QY 4 TIPLSRFDNLAHRLHQLAFDTYQFFEFAYIPKQKYSFLQNPQTSIFSSEIPTPS 63  
 DB 29 TVPLSRFDNLAHRLHQLAFDTYQFFEFAYIPKQKYSFLHDSOTSFCFSDSIPTPS 88  
 QY 64 NRETOQSNLELLRLISLLLIQSWLEPVQ 92



```

J. Biol. Chem. 260:9574-9579(1985).
[2]
SEQUENCE FROM N.A.
RP STRAIN-ETIDU: TISSUE=Liver;
RX MEDLINE-96194803; PubMed-8647448;
RA Das P., Meyer L., Seyfert H.-M., Brockmann G., Schwerin M.:
RT "Structure of the growth hormone-encoding gene and its promoter in
   mice.";
RL Gene 169:209-213(1996).
CC -1- FUNCTION: Plays an important role in growth control. Its major
   role in stimulating body growth is to stimulate the liver and
   other tissues to secrete IGF-1. It stimulates both the
   differentiation and proliferation of myoblasts. It also stimulates
   amino acid uptake and protein synthesis in muscle and other
   tissues.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation
   at the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X02891; CAA26550.1;
DR EMBL: Z46663; CAA86658.1;
DR PIR: B23911; STMS.
DR HSSP: P01246; LBST.
DR MGI: MGI:95707; Gh.
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone; 1.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
DR Hormone: Pituitary; Signal.
KW SIGNAL
FT CHAIN 1 26 BY SIMILARITY.
FT DISULFID 78 189 BY SIMILARITY.
FT DISULFID 206 214 BY SIMILARITY.
SQ SEQUENCE 216 AA: 24716 MW: 986666 JAE25D65FC CRC64:
Query Match 64.8%; Score 302.5; DB 1; Length 216;
Best Local Similarity 64.8%; Pred. No. 1, 20-24;
Matches 59; Conservative 14; Mismatches 17; Indels 1; Gaps 1;
QY 2 FPTIPLSRFDNAMLRAHLHQLAFDTYQFFEFAYIPKEQKYSFQNPQTSLSFSESIPT 61
DB 27 FPMPLSSLSFNAVLAQHLHQLAADTYKEFAYIPKEQKYS-IGNAQAACFCSETIPA 85
QY 62 PSNREETOQKSNLELLRISLLLIQSWLEPVQ
DB 86 PTGKDEAQAQRSDMELLRFSLLLIQLSWLGRVQ 6

```

```

*Sequence of a cDNA encoding horse growth hormone.*;
Gene 143:299-300(1994).
[2]
SEQUENCE OF 27-216.
RX MEDLINE-77005410; PubMed-965151;
RA Zakin M.M., Poskus E., Langton A.A., Ferrara P., Santome J.A.,
   Dellacha J.M., Paladini A.C.;
RT "Primary structure of equine growth hormone.";
RL Int. J. Pept. Protein Res. 8:435-444(1976).
CC [3]
CC PRELIMINARY SEQUENCE OF 27-216.
RX MEDLINE-7402362; PubMed-4747849;
RA Zakin M.M., Poskus E., Dellacha J.M., Paladini A.C., Santome J.A.;
RT "The amino acid sequence of equine growth hormone.";
RL FEBS Lett. 34:353-355(1973).
CC [4]
CC SEQUENCE OF 68 95 AND 183-216.
RA Zakin M.M., Poskus E., Dellacha J.M., Paladini A.C., Santome J.A.;
RT "Amino acid sequences around the cysteine residues in equine growth
   hormone.";
RL FEBS Lett. 25:77-82(1972).
CC [5]
CC SEQUENCE OF 202-216.
RX MEDLINE-68368390; PubMed-4876100;
RA Oliver L., Hartree A.S.;
RT "Amino acid sequences around the cysteine residues in horse growth
   hormone.";
RL Biochem. J. 109:19-24(1968).
CC -1- FUNCTION: Plays an important role in growth control. Its major
   role in stimulating body growth is to stimulate the liver and
   other tissues to secrete IGF-1. It stimulates both the
   differentiation and proliferation of myoblasts. It also stimulates
   amino acid uptake and protein synthesis in muscle and other
   tissues.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation
   at the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U02929; AAA21027.1;
DR HSSP: P01246; LBST.
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone; 1.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
DR Hormone: Pituitary; Signal.
KW SIGNAL
FT CHAIN 1 26 BY SIMILARITY.
FT DISULFID 78 189 BY SIMILARITY.
FT DISULFID 206 214 BY SIMILARITY.
SQ SEQUENCE 216 AA: 24423 MW: 97AB3173834E11AC CRC64:
Query Match 64.4%; Score 302.5; DB 1; Length 216;
Best Local Similarity 64.8%; Pred. No. 20-24;
Matches 59; Conservative 14; Mismatches 17; Indels 1; Gaps 1;
QY 2 FPTIPLSRFDNAMLRAHLHQLAFDTYQFFEFAYIPKEQKYSFQNPQTSLSFSESIPT 61
DB 27 FPMPLSSLSFNAVLAQHLHQLAADTYKEFAYIPKEQKYS-IGNAQAACFCSETIPA 85
QY 62 PSNREETOQKSNLELLRISLLLIQSWLEPVQ 92
DB 86 PTGKDEAQAQRSDMELLRFSLLLIQLSWLGRVQ 116

```

RESULT 14  
SOMA\_RABIT







## RESULT 2

ID 014644 PRELIMINARY: PRT: 245 AA  
 AC 014644:  
 DT 01-JAN-1998 (TREMBlrel: 05, Created)  
 DT 01-JAN-1998 (TREMBlrel: 05, last sequence update)  
 DT 01-DEC-2001 (TREMBlrel: 19, last annotation update)  
 DE Placental growth hormone isoform hGH-V3 precursor.  
 GN hGH-V.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Term placenta;  
 RX MEDLINE=98373737; PubMed=9709963;  
 RA Bouzewska C.J., Svensson P.A., Carlsson T., Clark R.,  
 Carlsson L.M.S., Carlsson B.  
 RT \*Cloning of two novel growth hormone transcripts expressed in human  
 placenta.\*  
 RL J. Clin. Endocrinol. Metab. 83:2878-2885(1998).  
 DR EMBL: AF060661; AAB71829.1;  
 DR HSSP: P01241; IA22.  
 DR InterPro: IPR001400; Somatotropin.  
 DR Pfam: PF00103; hormone: 1.  
 DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
 KW Signal.  
 FT SIGNAL: 1 26 POTENTIAL.  
 SQ SEQUENCE 245 AA: 27101 MW: 140078CD74 91CB CRC64;  
 Query Match 89.8%; Score 422; ID 4; Length 245;  
 Best Local Similarity 92.3%; Pred. No. 1 245-39;  
 Matches 84; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 2 FTPIPLSRFDNAMLRAHRLHQLAFDTYOEFEEAYIPKQKYSFLONPOTSISFSESIPT 61

Db 27 FTPIPLSRFDNAMLRAHRLHQLAFDTYOEFEEAYIPKQKYSFLONPOTSISFSESIPT 86

Qy 62 PSNRETOQKSNLELLRLISLLIQSWLEPVQ 92

Db 87 PSNRETOQKSNLELLRLISLLIQSWLEPVQ 117

## RESULT 3

ID 007369 PRELIMINARY: PRT: 217 AA.  
 AC 007369:  
 DT 01-NOV-1996 (TREMBlrel: 01, Created)  
 DT 01-NOV-1996 (TREMBlrel: 01, last sequence update)  
 DT 01-MAR-2003 (TREMBlrel: 23, last annotation update)  
 DE Chorionic somatomotropin-3.  
 OS Macaca mulatta (Rhesus macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciae;  
 OC Cercopitheciae; Macaca.  
 OX NCBI\_TaxID=9544;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Midpregnancy placenta;  
 RX MEDLINE=94008724; PubMed=8404617;  
 RA Golos T.G., Durnung M., Fisher J.M., Fowler P.D.;  
 RT \*Cloning of four growth hormone/chorionic somatomotropin-related  
 complementary deoxyribonucleic acids differentially expressed during  
 pregnancy in the rhesus monkey placenta.\*  
 RL Endocrinology 133:1744-1752(1993).  
 DR EMBL: L16554; AAA1884.1;  
 DR HSSP: P01241; IAXI.  
 DR InterPro: IPR001400; Somatotropin.  
 DR Pfam: PF00103; hormone: 1.  
 DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
 DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
 SQ SEQUENCE 217 AA: 24874 MW: 11686AFDBA1B185 CRC64;

Query Match 84.5%; Score 397; DB 6; Length 217;  
 Best Local Similarity 83.3%; Pred. No. 6 9e-37;  
 Matches 75; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

Qy 3 PTIPLSRFDNAMLRAHRLHQLAFDTYOEFEEAYIPKQKYSFLONPOTSISFSESIPT 62  
 Db 26 PSVPLSRFDNAMLRAHRLHQLAFDTYOEFEEAYIPKQKYSFLONPOTSISFSESIPT 87

Qy 63 SNRETOQKSNLELLRLISLLIQSWLEPVQ 92  
 Db 88 SNRETOQKSNLELLRLISLLIQSWLEPVQ 117

RESULT 4  
 Q07368  
 ID 007368 PRELIMINARY: PRT: 212 AA.  
 AC 007368:  
 DT 01-NOV-1996 (TREMBlrel: 01, Created)  
 DT 01-NOV-1996 (TREMBlrel: 01, last sequence update)  
 DT 01-MAR-2003 (TREMBlrel: 23, last annotation update)  
 LE Somatotropin 2 precursor (Growth hormone 2) (Fragment).  
 OS Macaca mulatta (Rhesus macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciae;  
 OC Cercopitheciae; Macaca.  
 OX NCBI\_TaxID=9544;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=94008724; PubMed=8404617;  
 RA Golos T.G., Durnung M., Fisher J.M., Fowler P.D.;  
 RT \*Cloning of four growth hormone/chorionic somatomotropin-related  
 complementary deoxyribonucleic acids differentially expressed during  
 pregnancy in the rhesus monkey placenta.\*  
 RL Endocrinology 133:1744-1752(1993).  
 DR EMBL: L16553; AAA1884.1;  
 DR HSSP: P01241; IAXI.  
 DR InterPro: IPR001400; Somatotropin.  
 DR Pfam: PF00103; hormone: 1.  
 DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 212 AA: 24525 MW: 27BC911062566F5 CRC64;  
 Query Match 84.3%; Score 396; DB 6; Length 212;  
 Best Local Similarity 82.2%; Pred. No. 8 7e-17;  
 Matches 74; Conservative 11; Mismatches 5; Indels 0; Gaps 0;

Qy 3 PTIPLSRFDNAMLRAHRLHQLAFDTYOEFEEAYIPKQKYSFLONPOTSISFSESIPT 62  
 Db 23 PSVPLSRFDNAMLRAHRLHQLAFDTYOEFEEAYIPKQKYSFLONPOTSISFSESIPT 82

Qy 63 SNRETOQKSNLELLRLISLLIQSWLEPVQ 92  
 Db 83 SNLETOQKSNLELLRLISLLIQSWLEPVQ 112

RESULT 5  
 Q07367  
 ID 007367 PRELIMINARY: PRT: 217 AA.  
 AC 007367:  
 DT 01-NOV-1996 (TREMBlrel: 01, Created)  
 DT 01-NOV-1996 (TREMBlrel: 01, last sequence update)  
 DT 01-MAR-2003 (TREMBlrel: 23, last annotation update)  
 DE Chorionic somatomotropin-1.  
 OS Macaca mulatta (Rhesus macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciae;  
 OC Cercopitheciae; Macaca.  
 OX NCBI\_TaxID=9544;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Midpregnancy placenta;

```

DR PROSITE: PS00338; SOMATOTROPIN_2;1
SQ SEQUENCE 217 AA; 24994 MW; 39FAACDDH6B2E951 CRC64;

Query Match
Best Local Similarity 81.1%; Score 381; DB 4; Length 217;
Matches 73; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

QY 4 TPLSLRFLDNLMLRAHLRLHQLAFYTYQFEFAYIPKQKYSFLQNPQTSLSFSSESISPTPS 63
Db TVPLSLRFLDNLMLRAHLRLHQLAFYTYQFEFAYIPKQKYSFLQNPQTSLSFSSESISPTPS 63
QY 29 TVPLSLRFLDNLMLRAHLRLHQLAFYTYQFEFAYIPKQKYSFLQNPQTSLSFSSESISPTPS 88
Db TVPLSLRFLDNLMLRAHLRLHQLAFYTYQFEFAYIPKQKYSFLQNPQTSLSFSSESISPTPS 88

QY 64 NREETQOKSNLELRISLLLIQSWLEPVQ 92
Db NREETQOKSNLELRISLLLIQSWLEPVQ 92

QY 89 NMEETQOKSNLELRISLLLIQSWLEPVQ 117
Db NMEETQOKSNLELRISLLLIQSWLEPVQ 117

RESULT 7
CHUNK# PRELIMINARY; FRT: 217 AA
AC QKN:
DT 01-MAR-2002 (TFMBRel: 20, Created)
DI 01-MAR-2002 (TFMBRel: 20, Last sequence update)
DT 01-OCT-2002 (TFMBRel: 22, Last annotation update)
DE Growth hormone.
GN GH-V.
OS Ateles geoffroyi (Black-handed spider monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Ateles.
OX NCBI_TaxID=9509;
RN [1]
SEQUENCE FROM N.A.
RP Revol A., Esquivel D., Santiago D., Barrera-Saldana H.;
KT "Independent duplication of the growth hormone gene in three
RT Anthropoidcan lineages."
BL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF374235; AA:72287.1; -.
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone.
DR PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
SQ SEQUENCE 217 AA; 25293 MW; 741745A1B75C053E CRC64;

Query Match
Best Local Similarity 75.8%; Score 348; DB 6; Length 217;
Matches 69; Conservative 8; Mismatches 14; Indels 0; Gaps 0;

QY 2 FPTPLSLRFLDNLMLRAHLRLHQLAFYTYQFEFAYIPKQKYSFLQNPQTSLSSESISPT 61
Db FPTPLSLRFLDNLMLRAHLRLHQLAFYTYQFEFAYIPKQKYSFLQNPQTSLSSESISPT 61

QY 27 FPLPLSLRFLDNLMLRAHLRLHQLAFYTYQFEFAYIPKQKYSFLQNPQTSLSSESISPT 86
Db FPLPLSLRFLDNLMLRAHLRLHQLAFYTYQFEFAYIPKQKYSFLQNPQTSLSSESISPT 86

QY 62 PSNREETQOKSNLELRISLLLIQSWLEPVQ 92
Db PSNREETQOKSNLELRISLLLIQSWLEPVQ 92

QY 87 PNKEEVIKAKSSLELLHLSLLLIQSWLEPVQ 117
Db PNKEEVIKAKSSLELLHLSLLLIQSWLEPVQ 117

RESULT 8
CHUNK# PRELIMINARY; FRT: 202 AA
AC Q14643;
DI 01-JAN-1998 (TFMBRel: 05, Created)
DT 01-JAN-1998 (TFMBRel: 05, Last sequence update)
DT 01-JUN-2002 (TFMBRel: 21, Last annotation update)
DE Placental growth hormone 20kDa isoform precursor.
GN HGH-V.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RP TISSUE-Term placenta;
RX MEDLINE=98373737; PubMed=9709963;

```



```

SQ SEQUENCE 216 AA: 24509 MW: 1EC467A84CCFEB02 CRC64:
Query Match 64.1%; Score 301.5; DB 6; Length 216;
Best Local Similarity 64.8%; Pred. No. 4.3e-26;
Matches 59; Conservative 14; Mismatches 17; Indels 1; Gaps 1;

QY 2 FPTPLSLRFLFNAMLRHRLHQLAFDTYGEFEAYIPKEQKYSFLONPOTSLSFSESIPT 61
DB 27 FPAMPSSLSFANAVLRAOHLHQLAADTYKEFERTYIPKEQKYS-IONQAACFSEETIPA 85
QY 62 PSNREETOQKSNLELLRISLLLIQSMLGPVQ 92
DB 86 PTGKFAQORSVDVLLRFSLLLIQSMLGPVQ 116

RESULT 12
O8HYE5
ID Q8HYE5 PRELIMINARY: PRT: 216 AA
AC Q8HYE5
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Growth hormone precursor
OS Alluropoda melanoleuca (Giant panda).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Ursidae; Alluropoda.
OX NCBI_TaxID:9646;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE: Pituitary;
RA Liao M., Zhu M., Zhang A.;
RT "Cloning and expression of cDNA encoding growth hormone from
RT Alluropoda melanoleuca."
RL Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL: AF540536; AAN77228.1;
KW Signal.
FT SIGNAL.
FT SIGNAL.
SQ SEQUENCE 216 AA: 24383 MW: 44EC17EC44HCB056 CRC64:
Query Match 64.1%; Score 301.5; DB 6; Length 216;
Best Local Similarity 64.8%; Pred. No. 4.3e-26;
Matches 59; Conservative 14; Mismatches 17; Indels 1; Gaps 1;

QY 2 FPTPLSLRFLFNAMLRHRLHQLAFDTYGEFEAYIPKEQKYSFLONPOTSLSFSESIPT 61
DB 27 FPAMPSSLSFANAVLRAOHLHQLAADTYKEFERTYIPKEQKYS-IONQAACFSEETIPA 85
QY 62 PSNREETOQKSNLELLRISLLLIQSMLGPVQ 92
DB 86 PTGKFAQORSVDVLLRFSLLLIQSMLGPVQ 116

RESULT 13
Q9R2C3
ID Q9R2C3 PRELIMINARY: PRT: 216 AA
AC Q9R2C3
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 23, Last annotation update)
DE Growth hormone.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID:10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Nguyen T.N.K., Liebhaber S.A.;
RT "Mouse Growth Hormone Locus: Nucleotide Sequence and Phylogenetic
RT Analyses."
RL Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
DR EMBL: U34362; AAC99988.1;
DR HSSP: P01241; IAXI.
DR InterPro: IPR001400; Somatotropin.

```

```

DR Pfam: PF00103; hormone; 1.
DR PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
SQ SEQUENCE 216 AA: 24682 MW: EC2A06DA02536H18 CRC64:
Query Match 63.5%; Score 296.5; DB 11; Length 216;
Best Local Similarity 63.7%; Pred. No. 9.4e-26;
Matches 58; Conservative 14; Mismatches 18; Indels 1; Gaps 1;

QY 2 FPTPLSLRFLFNAMLRHRLHQLAFDTYGEFEAYIPKEQKYSFLONPOTSLSFSESIPT 61
DB 27 FPAMPSSLSFANAVLRAOHLHQLAADTYKEFERTYIPKEQKYS-IONQAACFSEETIPA 85
QY 62 PSNREETOQKSNLELLRISLLLIQSMLGPVQ 92
DB 86 PTGKFAQORSVDVLLRFSLLLIQSMLGPVQ 116

RESULT 14
Q9J4M4
ID Q9J4M4 PRELIMINARY: PRT: 204 AA
AC Q9J4M4
DT 01-FEB-1997 (TREMBlrel. 62, Created)
DT 01-FEB-1997 (TREMBlrel. 62, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 1; 23, Last annotation update)
DE Placental growth hormone (Fragment).
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID:9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE: Placenta;
RA Lacroix M.C., Devincny E., Servely J.L., Puissant C., Kann G.;
RT "Expression of the growth hormone gene in ovine placenta: detection
RT and cellular localization of the protein."
RL Endocrinology 137:4886-4892(1996).
DR EMBL: U49063; AAC48679.1;
DR HSSP: P01241; IAXI.
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone; 1.
DR PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
DR NON_TER.
FT NON_TER.
SQ SEQUENCE 204 AA: 23462 MW: EBBB9451942635C5 CRC64:
Query Match 63.3%; Score 297.5; DB 6; Length 204;
Best Local Similarity 64.8%; Pred. No. 3.3e-25;
Matches 59; Conservative 14; Mismatches 17; Indels 1; Gaps 1;

QY 2 FPTPLSLRFLFNAMLRHRLHQLAFDTYGEFEAYIPKEQKYSFLONPOTSLSFSESIPT 61
DB 15 FPAMPSSLSFANAVLRAOHLHQLAADTYKEFERTYIPKEQKYS-IONQAACFSEETIPA 73
QY 62 PSNREETOQKSNLELLRISLLLIQSMLGPVQ 92
DB 74 PTGKFAQORSVDVLLRFSLLLIQSMLGPVQ 104

RESULT 15
Q9J4M4
ID Q9J4M4 PRELIMINARY: PRT: 216 AA
AC Q9J4M4
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Growth hormone precursor.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

OC Mammalia: Eutheria: Rodentia: Hystricognathi: Caviidae: Cavia.  
 OX NCBI\_TaxID:10141;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Odorico D.M., Fuller P.J., Herington A.C.  
 RT "Cloning and sequence of guinea pig growth hormone (GH).";  
 RI Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF233853; AAF36409.1;  
 DR HSSP: P01241; IAXI.  
 DR InterPro: IPR001400; Somatotropin.  
 DR Pfam: PF00103; hormone; 1.  
 DR PRINTS: PR00836; SOMATOTROPIN.  
 DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
 DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
 KW Signal.  
 FT SIGNAL 1 26 POTENTIAL.  
 FT CHAIN 27 216 GROWTH HORMONE.  
 SQ SEQUENCE 216 AA: 24622 MW: 45560pI: 46000.03664;

Query Match 64.8% Score 26.0 E-16.111 Length 216.  
 Best Local Similarity 64.7% Posit. Pos. 1 26 27  
 Matches 58: Conservative 14: Mismatches 18: Indels 1: Gaps 1:  
 QY 2 FVTPLSLRFDNMLRAHRLHOLAFDTYGFEEAVIPKEQKYSFLONPQTSLSFSESPT 61  
 II III II III III III III III III III III III III III III III  
 DB 27 FFAFPLSSLFQNAVLRRAHRLHOLAAATYKLRTYIPESQRYSLHNTQTAFCESETIPA 85  
 QY 62 PSNRFTQKSNLELLRLISLLLIQSWLPPVQ 92  
 III III III III III III III III III III III III III III III  
 DB 86 PTKKEFAQQRSDVELHFSLLLIQSWLPPVQ 11

Search completed: September 15, 2003, 12:04:27  
 Job time: 35.9427 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:44:15 : Search time 60.4158 Seconds  
(without alignments)  
225,942 Million cell updates/sec

Title: US-09-423-100-4

Perfect score: 463

Sequence: 1 FVQHLGCGSHVEALYVCG.....IVEQCCTICSLEYENVCN 86

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1197863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 117366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*

1:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Find. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	463	100.0	86	5	AAP40829
2	463	100.0	86	17	AAR84061
3	463	100.0	86	20	AAV42858
4	463	100.0	86	21	AA812770
5	463	100.0	86	23	AAH48218
6	463	100.0	87	3	AAP20036
7	463	100.0	87	5	AAP40217
8	463	100.0	87	6	AAP50127
9	463	100.0	87	6	AAP50060

10	463	100.0	87	7	AAP61090	Sequence encoded b
11	463	100.0	87	14	AAR32367	Proinsulin protein
12	463	100.0	88	11	AAR07682	Modified human ins
13	463	100.0	88	14	AAR33855	hpt. Homo sapiens
14	463	100.0	92	13	AAR20467	Yeast alpha-factor
15	463	100.0	93	10	AAP90102	Synthetic proinsul
16	463	100.0	96	20	AAO17830	Human proinsulin *
17	463	100.0	96	20	AAO17830	Human proinsulin p
18	463	100.0	97	15	AAR68898	Human pro-insulin
19	463	100.0	97	21	AA812773	Human proinsulin p
20	463	100.0	110	2	AAP10042	Sequence encoded b
21	463	100.0	110	2	AAP10053	Sequence of prepro
22	463	100.0	110	5	AAP40309	Sequence of human
23	463	100.0	110	20	AAV06608	Human preproinsuli
24	463	100.0	110	21	AAB26765	Human preproinsuli
25	463	100.0	110	21	AAB06144	Human insulin. Ho
26	463	100.0	110	21	AAV70366	Human Insulin used
27	463	100.0	110	21	AAV44367	Human proinsulin.
28	463	100.0	110	22	AAV65677	Human proinsulin n
29	463	100.0	110	22	AAE10337	Human pre proinsul
30	463	100.0	110	22	AA845424	Secretory cell lin
31	463	100.0	110	23	AA841590	Human preproinsuli
32	463	100.0	110	23	AA860634	Human insulin. Ho
33	463	100.0	117	21	AAV69788	MWpsp-MWmp10-Met-
34	463	100.0	130	21	AAV65787	MWpsp-MWmp10-(His
35	463	100.0	51	20	AAW81856	Human proinsulin a
36	463	100.0	53	21	AAV53589	Human preproinsuli
37	463	100.0	51	20	AAW81857	Human proinsulin a
38	463	100.0	51	20	AAW81857	Human proinsulin a
39	463	100.0	51	20	AAW81857	Human proinsulin a
40	463	100.0	51	20	AAW81857	Human proinsulin a
41	463	100.0	51	20	AAW81857	Human proinsulin a
42	463	100.0	51	20	AAW81857	Human proinsulin a
43	463	100.0	51	20	AAW81857	Human proinsulin a
44	463	100.0	51	20	AAW81857	Human proinsulin a
45	463	100.0	51	20	AAW81857	Human proinsulin a

#### ALIGNMENTS

RESULT 1  
AAP40829  
ID AAP40829 standard; Protein; 86 AA.  
XX  
AC AAP40829;  
XX  
DI 25-MAR-2003 (updated)  
DI 03-AUG-1992 (first entry)  
XX  
DE Sequence of human insulin precursor.  
XX  
KW Insulin precursor; connecting peptide; diabetes; hormone.  
XX  
QS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FH Region 1..30 /label= chain H  
FH Peptide 31..65 /label= connecting peptide  
FH Region 66..86 /label= chain A  
FH Modified-site 1 /label= F-NH2-H  
FH /note= "H or a chemically or enzymatically cleavable AA residue or peptide residue"  
FT Disulfide bond 7..72  
FT Disulfide bond 19..85  
FT Disulfide bond 71..76  
FT Modified-site 86 /label= N-OH  
XX

PN US4430266-A.  
 XX 07-FEB-1984.  
 PD 16-FEB-1982; 82US-0349397.  
 XX 16-FEB-1982; 82US-0349397.  
 PR 27-MAR-1980; 80US-0134389.  
 PR 28-NOV-1980; 80US-0210696.  
 XX (ELLI.) LILLY & CO. ELI.  
 PA Frank BH;  
 PI WPI: 1984-049032/08.  
 DR Insulin precursor prodn. from linear S-sulphonate and mercaptan  
 XX in single step without separate oxidn.  
 PT Claim 17; Column 4; 8pp; English.  
 XX The inventors claim a method for the prepn. of an insulin precursor  
 CC in which the A-chain and B-chain are joined through a connecting  
 CC peptide. The connecting peptide joins the A-chain at the amino  
 CC group of A-1 to the B-chain at the carboxyl group of B-30. The  
 CC method is pref. for the prepn. of human insulin precursor (see  
 CC AAP40829). The S0s of the connecting peptides of a number of species  
 CC are given (see AAP40828, AAP40830-39).  
 CC (Updated on 25-MAR-2003 to correct PA field.)  
 XX Sequence 86 AA:  
 SQ Query Match 100.0%; Score 463; Db 5; Length 86;  
 Best Local Similarity 100.0%; Pred. No. 8.4e-43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRRAEDLQVQLVQVGGPGAGSLQPLALEG 60  
 DB 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRRAEDLQVQLVQVGGPGAGSLQPLALEG 60  
 OY 61 SLQKRGIVEQCCTSIQSLYQLENYCN 86  
 DB 61 SLQKRGIVEQCCTSIQSLYQLENYCN 86  
 RESULT 2  
 ID AAR84061 standard; Protein; 86 AA.  
 XX AAR84061;  
 AC 22-AUG-1996 (first entry)  
 DT Human insulin.  
 DE Insulin; transformation; gene expression; fungi; fungal cell;  
 KW hormone; A-chain; C-chain; glycosylation.  
 KW Homo sapiens.  
 OS Key location/Qualifiers  
 XX CUS 1..261  
 FT /\*tag= a  
 FT /product= Insulin.  
 XX EP704527-A.  
 PN 03-APR-1996.  
 PD 03-AUG-1995; 95EP-0001999.  
 XX 05-AUG-1994; 94HR-0000432.  
 XX

PA (MEST/) MESTRIC S.  
 PA (PLIV.) PLIVA PHARM & CHEM FAH.  
 XX Mestric S., Punt P.J., Valinger K., Van DEN HONDEL CAMJQ;  
 XX WPI: 1996-129917/14.  
 DR N-PSDB; AAT17830, AAT17831.  
 XX DNA encoding human insulin precursors - which comprise B- and  
 PT A-chains linked via amino acid chain contg. 1 or more glycosylation  
 PT sites, for prepn. of insulin in fungal cells  
 XX Disclosure: Figure 1; 32pp; English.  
 XX DNA sequences encoding insulin precursors of formula R-Pq A, where B  
 CC and A represent B- and A-chains of insulin respectively, and Pq  
 CC represents a modified C-peptide or any number of amino acids  
 CC comprising at least one glycosylation consensus site, can be  
 CC inserted into expression vectors which in turn can be used to  
 CC transform fungal host cells. The fungal cells are then cultured  
 CC and the insulin expressed in such cells can be harvested.  
 XX Sequence 86 AA:  
 SQ Query Match 100.0%; Score 463; Db 17; Length 86;  
 Best Local Similarity 100.0%; Pred. No. 8.4e-43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRRAEDLQVQLVQVGGPGAGSLQPLALEG 60  
 DB 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRRAEDLQVQLVQVGGPGAGSLQPLALEG 60  
 OY 61 SLQKRGIVEQCCTSIQSLYQLENYCN 86  
 DB 61 SLQKRGIVEQCCTSIQSLYQLENYCN 86  
 RESULT 3  
 ID AAY42858 standard; Protein; 86 AA.  
 XX AAY42858;  
 AC 19-JAN-2000 (first entry)  
 DT Human insulin precursor, SEQ ID 5.  
 DE Insulin; precursor; growth hormone; chapetone; intramolecular;  
 KW folding; conformation; chimeric protein; clavable; recombinant;  
 KW production; yield.  
 XX Homo sapiens.  
 OS WO9950302-A1.  
 XX 07-OCT 1999.  
 PD 31-MAR 1998; 98WO-CN00052.  
 PF 31-MAR-1998; 98WO-CN00052.  
 PR (TONG.) TONGHUA GANTECH BIOTECHNOLOGY LTD.  
 PA Gan Z;  
 PI WPI: 1999-610839/52.  
 XX New chimeric proteins containing human growth hormone fragment, used  
 PT particularly for the production of human insulin -  
 XX Claim 10; Page 29; 46pp; English.  
 XX This sequence represents a human insulin precursor comprising insulin



XX WPI: 2002-082942/11.  
 XX Treating cartilage disorders including cartilage damage by injury or  
 XX PT degenerative cartilaginous disorders, by contacting cartilage with  
 XX PT insulin-like growth factor analog with altered affinity for IGF-binding  
 XX PT proteins.  
 XX PS Disclosure: Fig 16; 136pp; English.  
 XX The present invention relates to a method for treating cartilage  
 XX disorders. The method comprises contacting cartilage with an active agent  
 XX CC such as insulin-like growth factor (IGF 1) analog with a binding affinity  
 XX CC preference for IGF binding protein-3 (IGFBP-3) over IGFBP-1, an IGF-1  
 XX CC analog with a binding affinity preference for IGFBP-1 over IGFBP-3, or a  
 XX CC IGFBP displacer peptide that prevents the interaction of IGF with an  
 XX CC IGFBP and does not bind to human IGF receptor. The method is useful for  
 XX CC treating cartilage disorders (CD), including degenerative CD, articular  
 XX CC CD such as rheumatoid arthritis and osteoarthritis. The present sequence  
 XX CC is human proinsulin, which was used to illustrate the invention.  
 XX SQ Sequence 86 AA:

Query Match 100.0%; Score 463; DB 23; Length 86;  
 Best Local Similarity 100.0%; Pred. No. 4e-43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FVNORHCGSHLVEALYLVCGERGFFYTPKTRKEAEDLVQGVDELGGPGAGSLQPLALEG 60  
 DB 1 FVNORHCGSHLVEALYLVCGERGFFYTPKTRKEAEDLVQGVDELGGPGAGSLQPLALEG 60  
 QY 61 SLORRGIVEQCCTSCSLYOLENYCN 86  
 DB 61 SLORRGIVEQCCTSCSLYOLENYCN 86

RESULT 6  
 AAP20036  
 ID AAP20036 standard; Protein: 87 AA.

XX AC AAP20036;  
 XX DT 25-MAR-2003 (updated)  
 XX DT 22 JUL-1992 (first entry)  
 XX DE Human proinsulin.  
 XX KW Proinsulin.  
 XX OS Homo sapiens.  
 XX PA EP55942-A.  
 XX PD 14 JUL 1982.  
 XX PF 31 DEC 1981; 81EP-0306190.  
 XX PR 02 JAN 1981; 81US-0222010.  
 XX PR 23 JUL 1981; 81US-0286070.  
 XX PR 02 JAN 1982; 82US-0222010.  
 XX PR 03 MAR 1982; 82US-0354287.  
 XX PA (CNY) STATE UNIV NEW YORK.  
 XX PI Inouye M. Nakamura K;  
 XX WPI: 1982-59775E/29 (59775E).  
 XX DR N FSDB; AAN20041.  
 XX The present invention relates to a method for transforming bacterial  
 XX PT hosts to produce eukaryotic polypeptide(s).  
 XX PS Disclosure: Fig 27; 114pp; English.

XX CC The sequence comprises human proinsulin.  
 XX CC (Updated on 25-MAR-2003 to correct PR field.)  
 XX SQ Sequence 87 AA:  
 Query Match 100.0%; Score 463; DB 43; Length 87;  
 Best Local Similarity 100.0%; Pred. No. 4e-43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FVNORHCGSHLVEALYLVCGERGFFYTPKTRKEAEDLVQGVDELGGPGAGSLQPLALEG 60  
 DB 2 FVNORHCGSHLVEALYLVCGERGFFYTPKTRKEAEDLVQGVDELGGPGAGSLQPLALEG 61  
 QY 61 SLORRGIVEQCCTSCSLYOLENYCN 87  
 DB 62 SLORRGIVEQCCTSCSLYOLENYCN 87  
 RESULT 7  
 AAP40217  
 ID AAP40217 standard; Protein: 87 AA.  
 XX AC AAP40217;  
 XX DT 25-MAR-2003 (updated)  
 XX DT 12 FEB-1992 (first entry)  
 XX DE Sequence of the 32 N terminal AAs of proinsulin.  
 XX KW Hormone; cloning vector; phage resistant.  
 XX OS Homo sapiens.  
 XX FH Key Location/Qualifiers  
 XX FT Region 2...31  
 XX FT /label= B-chain  
 XX FT Region 32...66  
 XX FT /label= C-chain  
 XX FT Region 67...87  
 XX FT /label= A-chain  
 XX PN GB2126237-A.  
 XX PD 21-MAR 1984.  
 XX PF 01-SEP 1983; 83GB-0023468.  
 XX PR 03-SEP-1982; 82US-0414290.  
 XX PR 05-SEP 1984; 84US-0647338.  
 XX PA (EHL) LILLY & CO ELI.  
 XX PI Hershey CL. Rostock PR;  
 XX DR WPI: 1984-070793/12.  
 XX DR N-PSDB; AAN40179.  
 XX PT Protecting bacteria from phage infection - by transformation with  
 XX PT cloning vector contg. segment with restriction and modification  
 XX PT activity  
 XX PS Example: Fig 10; 28pp; English.  
 XX CC Plasmid pTh alpha 1 was constructed by inserting a synthesised gene  
 XX CC for thymosin alpha 1 (AAN40178) into plasmid pBR322. It is used for  
 XX CC the construction of pTrp24. The inventors claim a method for  
 XX CC protecting bacteria from phage infection - by transformation with  
 XX CC cloning vector contg. segment with restriction and modification  
 XX CC activity. Prodn. of plasmid pBR 26 or pBR27 which uses pTrp24; and  
 XX CC prodn. of plasmid pBR29 which uses a synthetic gene coding for the  
 XX CC 32 N-terminal AAs of proinsulin (see AAN40179).  
 XX CC (Updated on 25-MAR-2003 to correct PA field.)



CC readily isolated from cell lysates and is stable on washing with  
 CC urea or detergent solns. at low concns. The granule contains at  
 CC least 50% of proinsulin and all isolation operations are  
 CC simplified.

CC (Updated on 16-AUG-2002 to add missing OS field.)  
 CC (Updated on 25-MAR-2003 to correct PA field.)

XX  
 XX Sequence 87 AA;

Query Match 100.0%; Score 463; DB 6; Length 87;  
 Best Local Similarity 100.0%; Pred. No. 8.5e-43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLVQGVFLGGPGAGSLQPLALEG 60  
 DB 2 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLVQGVFLGGPGAGSLQPLALEG 61  
 OY 61 SLOKRGIVEOCTTSCSLYLENYCN 86  
 DB 62 SLOKRGIVEOCTTSCSLYLENYCN 87

RESULT 10  
 AAP61090  
 ID AAP61090 standard; Protein: 87 AA.

XX  
 XX AAB61090;

XX  
 XX 28-FEB-1992 (first entry)

XX  
 XX Sequence encoded by the structural gene for human proinsulin.

XX  
 XX Recombinant plasmid; E. coli expression vector; secretion vector.

XX  
 XX Homo sapiens.

XX  
 XX US4624926-A.

XX  
 XX 25-Nov 1986.

XX  
 XX 03-MAR-1982; 82US-0354287.

XX  
 XX 03-MAR-1982; 82US-0354287.

XX  
 XX 23-JAN-1981; 81US-0222010.

XX  
 XX 23-JUL-1981; 81US-0286070.

XX  
 XX (UYNV-) UNIV OF NEW YORK.

XX  
 XX Inouye M, Nakamura K;

XX  
 XX WPI: 1986 331802/50.

XX  
 XX N-PSDB; AAN60872.

XX  
 XX New recombinant plasmid(s) - contg. DNA sequences encoding  
 XX exogenous polypeptide and outer membrane protein of E coli

XX  
 XX Example: Fig 27; 44pp; English.

XX  
 XX The inventors claim new recombinant plasmids contg. a DNA sequence  
 XX encoding a polypeptide, which is foreign to E.coli, in reading phase  
 XX with a DNA SQ, coding for at least one functional fragment derived  
 XX from an outer membrane lipoprotein gene of E.coli. The foreign gene  
 XX may be for human insulin. The lipoprotein gene functional fragment  
 XX may be the promoter, the 5'-UTR, the 3'-UTR or the transcription  
 XX termination signal provided that it includes at least the promoter.

XX Sequence 87 AA;

Query Match 100.0%; Score 463; DB 7; Length 87;  
 Best Local Similarity 100.0%; Pred. No. 8.5e-43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLVQGVFLGGPGAGSLQPLALEG 60

DB 2 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLVQGVFLGGPGAGSLQPLALEG 61  
 OY 61 SLOKRGIVEOCTTSCSLYLENYCN 86  
 DB 62 SLOKRGIVEOCTTSCSLYLENYCN 87

RESULT 11

AAR32367  
 ID AAR32367 standard; Protein: 87 AA.

XX  
 XX AAR32367;

XX  
 XX 25-MAR-2003 (updated)

XX  
 XX 18-JUN-1993 (first entry)

XX  
 XX Proinsulin protein sequence.

XX  
 XX Human; proinsulin; vector; pGCI9; pPINS; CAT; pHC-CAT proinsulin;

XX  
 XX Insulin; analogue; type I; type II; diabetes.

XX  
 XX Synthetic.

XX  
 XX W09303174-A1.

XX  
 XX 18-FEB-1993.

XX  
 XX 31-JUL-1992; 92WO-0506451.

XX  
 XX 08-AUG-1991; 91US-0741938.

XX  
 XX 30-JUL-1992; 92US-0518953.

XX  
 XX (PEIZ) PFIZER INC.

XX  
 XX (SCIO-) SCIOS INC.

XX  
 XX Andy R.J. Larson Esq.

XX  
 XX WPI: 1993-076530/09.

XX  
 XX N-PSDB; AAO37003.

XX  
 XX New hepato selective and peripheral selective human insulin

XX  
 XX analogues - and their corresp. DNA, for treatment of type I and

XX  
 XX type II diabetes

XX  
 XX Disclosure: Fig 2b; 88pp; English.

XX  
 XX This sequence represents human proinsulin and was decoded from  
 XX the sequences given in AAO36996-7001. The cDNA fragment coding for  
 XX proinsulin was inserted into plasmid vector pUC19 and digested with  
 XX KpnI and HindIII. This resulted in the formation of the vector pPINS.  
 XX A fragment encoding amino acids 1-73 of CAT (see AAO37002) was inserted  
 XX into pPINS to give a plasmid which contained DNA sequences which coded  
 XX for amino acids 1-73 of CAT, an 8 amino acid linker sequence and human  
 XX proinsulin. This plasmid, pHC-CAT-proinsulin, could be used in the  
 XX formation of insulin analogues which may be used in the treatment of  
 XX types I and II diabetes.

XX  
 XX (Updated on 25-MAR-2003 to correct PN field.)

XX  
 XX Sequence 87 AA;

Query Match 100.0%; Score 463; DB 14; Length 87;  
 Best Local Similarity 100.0%; Pred. No. 8.5e-43;

XX  
 XX Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLVQGVFLGGPGAGSLQPLALEG 60

DB 2 FVNHLCGSHLVEALYLV CGERGFFYTPKTRREAEDLVQGVFLGGPGAGSLQPLALEG 61

OY 61 SLOKRGIVEOCTTSCSLYLENYCN 86

DB 62 SLOKRGIVEOCTTSCSLYLENYCN 87

RESULT 12	AAAR07682	standard; protein: 88 AA.
ID	AAAR07682	standard; protein: 88 AA.
XX	AAAR07682	
XX	25-MAR-2003 (updated)	
XX	09-JAN-2003 (updated)	
DT	13-FEB-1991 (first entry)	
DT		
XX	Modified human insulin precursor.	
DE		
XX	Human insulin precursor; cathepsin C.	
KW		
KW	Homo sapiens.	
OS		
XX	Key	Location/Qualifiers
FT	Peptide	1..2
FT	Peptide	/label=N-terminal initiating dipeptide
FT	Peptide	3..32
FT	Peptide	/label=Native human insulin B chain
FT	Peptide	33..67
FT	Peptide	/label=natural connecting peptide of human proinsulin
FT	Peptide	68..88
FT	Peptide	/label=Native human insulin A-chain
XX	EP37420 A.	
PN		
XX	14-NOV-1990.	
XX	04-MAY-1990.	90EP-030489C.
XX	09-MAY-1989.	89US-0349472.
XX		
XX	(ELIL ) LILLY & CO ELI.	
PA		
XX	Recker GW, Furman TC, Mackellar WC, McDonough JP;	
PI	WPI: 1990-343372/46.	
DR		
XX	Human insulin precursor - contg. Met-Tyr or Met-arg initiating dipeptide for controlled removal by cathepsin C	
PT		
PT		
XX	Disclosure: page 3: 8pp: English.	
PS		
XX	This modified human insulin precursor comprises an N-terminal initiating dipeptide, chosen from Met-Tyr or Met-Arg, which does not define a cathepsin C dipeptide removal stop point. This dipeptide is linked to the natural human insulin B-chain, natural human proinsulin connecting peptide and natural human insulin A-chain. Dipeptide removal is carefully controlled to obtain the desired prod. without further degradation occurring, irrespective of whether the next dipeptide in the sequence defines a cathepsin C stop point.	
CC	(updated on 09-JAN-2003 to add missing (S field.)	
CC	(updated on 25-MAR-2003 to correct PA field.)	
XX		
XX	Sequence 88 AA:	
XX	Query Match 100.0%; Score 463; DR 11; Length 88;	
XX	Best Local Similarity 100.0%; Pred. No. 8.6e-43;	
XX	Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 FVNHLCGSHLVEALYLVCGERGEFFYTPKTRREAEDLQVGQVELGGPAGSLQPLEAG 60	
DB	3 FVNHLCGSHLVEALYLVCGERGEFFYTPKTRREAEDLQVGQVELGGPAGSLQPLEAG 62	
QY	61 SLOKRGIVEOCCTSIQSLYLENYCN 86	
DB	63 SLOKRGIVEOCCTSIQSLYLENYCN 88	
XX		
XX	Sequence 88 AA:	
XX	Query Match 100.0%; Score 463; DR 14; Length 88;	
XX	Best Local Similarity 100.0%; Pred. No. 8.6e-43;	
XX	Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	3 FVNHLCGSHLVEALYLVCGERGEFFYTPKTRREAEDLQVGQVELGGPAGSLQPLEAG 60	
DB	61 SLOKRGIVEOCCTSIQSLYLENYCN 86	
QY	63 SLOKRGIVEOCCTSIQSLYLENYCN 88	
DB		
XX		
XX	Sequence 88 AA:	
XX	Query Match 100.0%; Score 463; DR 14; Length 88;	
XX	Best Local Similarity 100.0%; Pred. No. 8.6e-43;	
XX	Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	3 FVNHLCGSHLVEALYLVCGERGEFFYTPKTRREAEDLQVGQVELGGPAGSLQPLEAG 60	
DB	61 SLOKRGIVEOCCTSIQSLYLENYCN 86	
QY	63 SLOKRGIVEOCCTSIQSLYLENYCN 88	
DB		
XX		
XX	Sequence 88 AA:	
XX	Query Match 100.0%; Score 463; DR 14; Length 88;	
XX	Best Local Similarity 100.0%; Pred. No. 8.6e-43;	
XX	Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	3 FVNHLCGSHLVEALYLVCGERGEFFYTPKTRREAEDLQVGQVELGGPAGSLQPLEAG 60	
DB	61 SLOKRGIVEOCCTSIQSLYLENYCN 86	
QY	63 SLOKRGIVEOCCTSIQSLYLENYCN 88	
DB		
XX		
XX	Sequence 88 AA:	
XX	Query Match 100.0%; Score 463; DR 14; Length 88;	
XX	Best Local Similarity 100.0%; Pred. No. 8.6e-43;	
XX	Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	3 FVNHLCGSHLVEALYLVCGERGEFFYTPKTRREAEDLQVGQVELGGPAGSLQPLEAG 60	
DB	61 SLOKRGIVEOCCTSIQSLYLENYCN 86	
QY	63 SLOKRGIVEOCCTSIQSLYLENYCN 88	
DB		
XX		
XX	Sequence 88 AA:	
XX	Query Match 100.0%; Score 463; DR 14; Length 88;	
XX	Best Local Similarity 100.0%; Pred. No. 8.6e-43;	
XX	Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	3 FVNHLCGSHLVEALYLVCGERGEFFYTPKTRREAEDLQVGQVELGGPAGSLQPLEAG 60	
DB	61 SLOKRGIVEOCCTSIQSLYLENYCN 86	
QY	63 SLOKRGIVEOCCTSIQSLYLENYCN 88	
DB		
XX		
XX	Sequence 88 AA:	
XX	Query Match 100.0%; Score 463; DR 14; Length 88;	
XX	Best Local Similarity 100.0%; Pred. No. 8.6e-43;	
XX	Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	3 FVNHLCGSHLVEALYLVCGERGEFFYTPKTRREAEDLQVGQVELGGPAGSLQPLEAG 60	
DB	61 SLOKRGIVEOCCTSIQSLYLENYCN 86	
QY	63 SLOKRGIVEOCCTSIQSLYLENYCN 88	

DE Yeast alpha-factor signal-human proinsulin fusion product.  
 XX BCA-5; yeast preferred codons; post-translational processing;  
 KW insulin; endopeptidase.  
 XX Synthetic.

XX Key Location/Qualifiers  
 XX Cleavage site 6..7  
 FT /note= "signal-proinsulin junction"  
 FT Cleavage site 37..38

XX US5077204-A.

XX 31-DEC-1991.

XX 08-APR-1988; 88US-0183252.

XX 21-JUN-1984; 84US-0523306

XX 08-APR-1988; 88US-0183252

XX (REG ) UNIV CALIFORNIA.

XX Blake AJ, Blair LC, Julius D, Thornton CB;

XX WPI; 1992-032671/04.

XX N-PSDB; AAQ20543.

XX Novel DNA for endopeptidase produ. - useful for in vivo or in  
 PT vitro processing of poly:peptide(s)

XX Example 1; Fig 1; 16pp; English.

XX The fusion product is encoded by a synthetic sequence having at its  
 CC 5'-end a modification of the 3'-end of the naturally occurring  
 CC alpha-factor secretory leader and processing signal sequence, where  
 CC three Glu Ala pairs have been deleted. A plasmid containing the  
 CC synthetic proinsulin coding sequence was used to transform kex2-  
 CC mutant yeast strains in the presence or absence of the cloned KEX2  
 CC gene. Post-translational processing of pro-insulin into peptides  
 CC only occurred in yeast transformed to kex2 plus. See also AAQ20545.  
 CC (updated on 25-MAR-2003 to correct PA field.)

XX Sequence 92 AA;

Query Match 100.0%; Score 463; DB 13; Length 92;  
 Best Local Similarity 100.0%; Pred. No. 96 43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNHLCGSHLVVALYLVGGERGFFVTPKTRRAEDLVQGVFLGGPGAGSLQPLALEG 60  
 Db 7 FVNHLCGSHLVVALYLVGGERGFFVTPKTRRAEDLVQGVFLGGPGAGSLQPLALEG 66

Qy 61 SLOKRGIVEQCCTISICSLYLENYCN 86

Db 67 SLOKRGIVEQCCTISICSLYLENYCN 92

RESULT 15

AAP90102

XX AAP90102 standard; protein; 93 AA.

XX AAP90102;

XX 25-MAR-2003 (updated)

DT 01-NOV-1989 (first entry)

XX Synthetic proinsulin.

XX Proinsulin; synthetic gene; yeast.

XX Homo sapiens.

XX Key Location/Qualifiers  
 FT Peptide 8..37  
 FT Region 38..39 "proinsulin chain B"  
 FT Peptide 38..72 "dibasic processing site"  
 FT Region 71..72 "proinsulin chain C"  
 FT Peptide 73..93 "dibasic processing site"  
 FT /note= "proinsulin chain A"

XX EP324274-A.

XX 19-JUL-1989.

XX 23-DEC-1988; 88EP-0312306.

XX 30-DEC-1987; 87US-0196652.

XX (CHIR ) CHIRUKIN O-48P.

XX Tokatopoliso P;

XX WPI; 1989-208260/29.

XX N-PSDB; AAN90279.

XX Expression and secretion of heterologous proteins in yeast - using  
 PT truncated alpha-factor leader sequence contn. signal; peptide and  
 PT glycosylation site.

XX Disclosure; fig 1; 24pp; English.

XX The proinsulin is expressed in Saccharomyces cerevisiae AB103.1.  
 CC It is encoded by a synthetic gene (see AAN90279). The 3 consecutive  
 CC peptides are proinsulin chains B, C and A resp. The regions are  
 CC dibasic processing sites.  
 CC (updated on 25-MAR-2003 to correct P1 field.)

XX Sequence 93 AA;

Query Match 100.0%; Score 463; DB 10; Length 93;  
 Best Local Similarity 100.0%; Pred. No. 43 43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNHLCGSHLVVALYLVGGERGFFVTPKTRRAEDLVQGVFLGGPGAGSLQPLALEG 60  
 Db 8 FVNHLCGSHLVVALYLVGGERGFFVTPKTRRAEDLVQGVFLGGPGAGSLQPLALEG 67

Qy 61 SLOKRGIVEQCCTISICSLYLENYCN 86

Db 68 SLOKRGIVEQCCTISICSLYLENYCN 94

Search completed: September 15, 2003, 12:00:58

Job time : 61.4158 secs

Gencore version 5.1.6  
Copyright (C) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:57:40 : Search time 19.7276 Seconds  
(without alignments)  
184.449 Million cell updates/sec

Title: US-09-423-100-4

Perfect score: 463

Sequence: 1 F NOHLCGSHLVEALYLVCGR.....IVFOCTCSICSLYLENYCN 86

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 4210458 residues

Total number of hits satisfying the Seq. Gap. Cuts: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents-AA:\*

- 1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*
- 2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*
- 3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*
- 4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*
- 5: /cgn2\_6/ptodata/1/iaa/PTUS\_COMB.pep:\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY:

Result No.	Score	Query Match	Length	DB ID	Description
1	463	100.0	86	US-09-477-924-2	Sequence 2, Appl
2	463	100.0	86	US-09-723-981-2	Sequence 2, Appl
3	463	100.0	86	US-09-723-846-2	Sequence 2, Appl
4	463	100.0	86	US-09-878-801-1	Sequence 1, Appl
5	463	100.0	96	US-09-134-77-4	Sequence 4, Appl
6	463	100.0	96	US-09-386-7A-4	Sequence 4, Appl
7	463	100.0	97	US-08-160-A-4	Sequence 4, Appl
8	463	100.0	110	US-08-950-A-11	Sequence 11, Appl
9	463	100.0	110	US-08-584-3-2	Sequence 2, Appl
10	463	100.0	110	US-08-784-7-2	Sequence 2, Appl
11	463	100.0	110	US-08-785-7-2	Sequence 2, Appl
12	463	100.0	110	US-08-472-7-1	Sequence 2, Appl
13	463	100.0	110	US-09-185-6-2	Sequence 2, Appl
14	463	100.0	110	PCT-US95-08596-2	Sequence 2, Appl
15	463	100.0	117	US-09-280-030-63	Sequence 63, Appl
16	463	100.0	130	US-09-280-030-62	Sequence 62, Appl
17	463	100.0	151	US-08-508-664-15	Sequence 15, Appl
18	463	100.0	161	US-08-508-664-16	Sequence 16, Appl
19	463	100.0	167	US-07-918-753-8	Sequence 8, Appl
20	463	100.0	167	US-08-081-7-8	Sequence 8, Appl
21	457	98.7	96	US-09-134-7-5	Sequence 5, Appl
22	457	98.7	96	US-09-386-7A-5	Sequence 5, Appl
23	457	98.7	97	US-08-389-4-7-7	Sequence 7, Appl
24	456	98.5	90	US-08-030-731A-43	Sequence 43, Appl
25	446	96.3	97	US-09-099-7-6	Sequence 6, Appl
26	444	95.9	97	US-09-099-307-8	Sequence 8, Appl
27	443	95.7	110	US-08-589-028-4	Sequence 4, Appl

Sequence 4, Appl  
Sequence 4, Appl  
Sequence 7, Appl  
Sequence 11, Appl  
Sequence 7, Appl  
Sequence 1, Appl  
Sequence 3, Appl  
Sequence 6, Appl  
Sequence 16, Appl  
Sequence 2, Appl  
Sequence 8, Appl  
Sequence 5, Appl  
Sequence 7, Appl  
Sequence 43, Appl  
Sequence 15, Appl  
Sequence 9, Appl  
Sequence 4, Appl  
Sequence 4, Appl

28 443 95.7 110 3 US-08-784-582-4  
29 443 95.7 110 3 US-08-785-271-4  
30 440 95.0 97 3 US-09-099-307-7  
31 435 94.0 97 3 US-09-099-307-11  
32 398 86.0 91 4 US-09-676-787-7  
33 292.5 63.2 67 4 US-08-981-988A-1  
34 290.5 62.7 83 4 US-08-981-988A-3  
35 288.5 62.3 83 4 US-08-981-988A-6  
36 288.5 62.3 113 4 US-09-484-848-16  
37 286.5 61.9 67 4 US-08-981-988A-2  
38 286.5 61.7 83 4 US-08-981-988A-8  
39 284.5 61.4 67 4 US-08-981-988A-5  
40 281.5 60.8 67 4 US-08-981-988A-7  
41 281 60.7 86 4 US-09-201-227A-43  
42 280 60.5 112 4 US-09-484-848-15  
43 279.5 60.4 83 4 US-08-981-988A-9  
44 278.5 60.2 69 4 US-08-981-988A-4  
45 272.5 58.9 53 1 US-08-231-517-4

ALIGNMENTS

RESULT 1  
US-09-477-924-2  
: Sequence 2, Application: US/09477924  
: Patent No. 6403764  
: GENERAL INFORMATION:  
: APPLICANT: Dubaquitte, Yves  
: APPLICANT: Lomman, Henry  
: TITLE OF INVENTION: PROTEIN VARIANTS  
: FILE REFERENCE: P1712R1-1  
: CURRENT APPLICATION NUMBER: US/09/477,924  
: CURRENT FILING DATE: 2000-11-05  
: NUMBER OF SEQ ID NOS: 6  
: SEQ ID NO 2  
: LENGTH: 86  
: TYPE: PRT  
: ORGANISM: Homo sapiens  
US-09-477-924-2

Query Match 100.0% Score 463; DB 4; Length 86;  
Best Local Similarity 100.0% Pred. No. 1.5e-47;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNOHLCGSHLVEALYLVCGRGFFYTPKTRRFAELQVGVGGSGPGAGSLQPLALEG 60  
DB 1 FVNOHLCGSHLVEALYLVCGRGFFYTPKTRRFAELQVGVGGSGPGAGSLQPLALEG 60  
QY 6: SLQKRGIVEQCTCSICSLYLENYCN 86  
DB 61 SLQKRGIVEQCTCSICSLYLENYCN 86

RESULT 2  
US-09-723-981-2  
: Sequence 2, Application US/09723981  
: Patent No. 6506874  
: GENERAL INFORMATION:  
: APPLICANT: Dubaquitte, Yves  
: APPLICANT: Lomman, Henry  
: TITLE OF INVENTION: PROTEIN VARIANTS  
: FILE REFERENCE: P1712R1  
: CURRENT APPLICATION NUMBER: US/09/723,981  
: CURRENT FILING DATE: 2000-11-28  
: PRIOR APPLICATION NUMBER: 09/477,923  
: PRIOR FILING DATE: 2000-01-05  
: NUMBER OF SEQ ID NOS: 6  
: SEQ ID NO 2  
: LENGTH: 86  
: TYPE: PRT  
: ORGANISM: Homo sapiens  
US-09-723-981-2

```

Query Match      100.0%   Score 463; DB 4; Length 86;
Best Local Similarity 100.0%; Pred. No. 1.5e-47;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREAEDELQGVQVELGGPGAGSLOPLALEG 60
    |||||||
DB 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREAEDELQGVQVELGGPGAGSLOPLALEG 60
    |||||||

QY 61 SLOKRGIVEQCCTSCISLYOLENYCN 86
    |||||||
DB 61 SLOKRGIVEQCCTSCISLYOLENYCN 86
    |||||||

RESULT 3
US-09-723-896 2
: Sequence 2, Application US/09723896
: Patent No. 6504443
: GENERAL INFORMATION:
: APPLICANT: Dubaque, Yves
: APPLICANT: Loman, Henry
: TITLE OF INVENTION: PROTEIN VARIANTS
: FILE REFERENCE: P1712R1
: CURRENT APPLICATION NUMBER: US/09/723,896
: CURRENT FILING DATE: 2000-11-28
: PRIOR APPLICATION NUMBER: US/09/477,924
: PRIOR FILING DATE: 2000-01-05
: NUMBER OF SEQ ID NOS: 6
: SEQ ID NO: 2
: LENGTH: 86
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-723-896 2

Query Match      100.0%   Score 463; DB 4; Length 86;
Best Local Similarity 100.0%; Pred. No. 1.5e-47;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREAEDELQGVQVELGGPGAGSLOPLALEG 60
    |||||||
DB 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREAEDELQGVQVELGGPGAGSLOPLALEG 60
    |||||||

QY 61 SLOKRGIVEQCCTSCISLYOLENYCN 86
    |||||||
DB 61 SLOKRGIVEQCCTSCISLYOLENYCN 86
    |||||||

RESULT 4
US-09-878-380 1
: Sequence 1, Application US/09878380
: Patent No. 6534281
: GENERAL INFORMATION:
: APPLICANT: Fujire Bio Inc.
: APPLICANT: KITAJIMA, Sachiko
: APPLICANT: KURAMO, Yoshihiro
: APPLICANT: NAKATSURO, Kaoru
: APPLICANT: NISHIZONO, Isao
: TITLE OF INVENTION: Immunoassay For Measuring Human C-Peptide and Kit Therefor
: FILE REFERENCE: 0760-0291P
: CURRENT APPLICATION NUMBER: US/09/878,380
: CURRENT FILING DATE: 2001-06-12
: PRIOR APPLICATION NUMBER: JP 2000-174669;
: PRIOR FILING DATE: 2000-06-12
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO: 1
: LENGTH: 86
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-878-380-1

Query Match      100.0%   Score 463; DB 4; Length 86;
Best Local Similarity 100.0%; Pred. No. 1.5e-47;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREAEDELQGVQVELGGPGAGSLOPLALEG 60
    |||||||
DB 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREAEDELQGVQVELGGPGAGSLOPLALEG 60
    |||||||

QY 61 SLOKRGIVEQCCTSCISLYOLENYCN 86
    |||||||
DB 61 SLOKRGIVEQCCTSCISLYOLENYCN 86
    |||||||

RESULT 5
US-09-134-836 4
: Sequence 4, Application US/09134836
: Patent No. 5980048
: GENERAL INFORMATION:
: APPLICANT: Kubroder, Franz Josef
: APPLICANT: Keller, Reinhold
: TITLE OF INVENTION: Improved process for obtaining
: TITLE OF INVENTION: resulting precursors having correctly bonded cystine bridges
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Finnegan, Henderson, Farabee, Garrett &
: STREET: 1300 I Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-3415
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/134,836
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Leslie McDonnell
: REGISTRATION NUMBER: 34,872
: REFERENCE/DOCKET NUMBER: 02481-1600-00000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 408-4000
: TELEFAX: (202) 408-4400
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 96 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: ORIGINAL SOURCE:
: ORGANISM: Escherichia coli
: FEATURE:
: NAME/KEY: Protein
: LOCATION: 1..96
US-09-134-836-4

Query Match      100.0%   Score 463; DB 4; Length 96;
Best Local Similarity 100.0%; Pred. No. 1.8e-47;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREAEDELQGVQVELGGPGAGSLOPLALEG 60
    |||||||
DB 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREAEDELQGVQVELGGPGAGSLOPLALEG 70
    |||||||

QY 61 SLOKRGIVEQCCTSCISLYOLENYCN 96
    |||||||
DB 71 SLOKRGIVEQCCTSCISLYOLENYCN 96
    |||||||

RESULT 6

```

```

US-09-386-303A-4
: Sequence 4, Application US/09386303A
: Patent No. 6380355
: GENERAL INFORMATION:
: APPLICANT: Rudroder, Franz-Josef
: TITLE OF INVENTION: Improved process for obtaining
: insulin precursors having correctly bonded cysteine bridges
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Finnegan, Henderson, Farrelow, Garrett &
: Dunner
: STREET: 1300 I Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-3315
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC DOS/MS-DOS
: SOFTWARE: Patent Release 1.0, Version 1.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/386,303A
: FILING DATE: 31-Aug-1999
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 09/134,836
: FILING DATE: <Unknown>
: ATTORNEY/AGENT INFORMATION:
: NAME: Leslie McDonnell
: REGISTRATION NUMBER: 34,872
: REFERENCE/DOCKET NUMBER: 02481 1600-00000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 408-4000
: TELEFAX: (202) 408-4400
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 96 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: ORIGINAL SOURCE:
: ORGANISM: Escherichia coli
: FEATURE:
: NAME/KEY: Protein
: LOCATION: 1..96
: SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-386-303A-4
Query Match 100.0%; Score 463; DB 4; Length 96;
Best Local Similarity 100.0%; Pred. No. 1.8e-47;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGVPGAGSLQPLALEG 60
Db 11 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGVPGAGSLQPLALEG 70
QY 61 SLKRGIVEQCCTSCISLYOLENYCN 86
Db 71 SLKRGIVEQCCTSCISLYOLENYCN 96
RESULT 7
US-08-160-376A-4
: Sequence 4, Application US/08160376A
: Patent No. 5473049
: GENERAL INFORMATION:
: APPLICANT: Obermeier, Ranier
: APPLICANT: Gerli, Martin
: APPLICANT: Ludwig, Jurgen
: APPLICANT: Sabel, Walter

```

```

: TITLE OF INVENTION: Process For Obtaining Proinsulin
: TITLE OF INVENTION: Possessing Correctly Linked
: TITLE OF INVENTION: Cystine Bridges
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Kenneth A. Genoni, Esq.
: STREET: Rt. 202-206 No. 5473049th/P.O. Box 2500
: CITY: Somerville
: STATE: New Jersey
: COUNTRY: U.S.A.
: ZIP: 08876-1258
: COMPUTER READABLE FORM:
: MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
: COMPUTER: IBM 386
: OPERATING SYSTEM: WINDOWS 3.1
: SOFTWARE: WORDPERFECT 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/160,376A
: FILING DATE: December 1, 1993
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GE P 424,420.7
: FILING DATE: December 2, 1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Barba; a. V. Maurel, Esq.
: REGISTRATION NUMBER: 31,287
: REFERENCE/DOCKET NUMBER: HOE 92/F 384
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (908) 231-4079
: TELEFAX: (908) 231-2255
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 97 Amino Acids
: TYPE: Amino Acid (AA)
: TOPOLOGY: not relevant
US-08-160-376A-4
Query Match 100.0%; Score 463; DB 1; Length 97;
Best Local Similarity 100.0%; Pred. No. 1.8e-47;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGVPGAGSLQPLALEG 60
Db 12 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGVPGAGSLQPLALEG 71
QY 61 SLKRGIVEQCCTSCISLYOLENYCN 86
Db 72 SLKRGIVEQCCTSCISLYOLENYCN 97
RESULT 8
US-08-950-720A-11
: Sequence 11, Application US/08950720A
: Patent No. 6046028
: GENERAL INFORMATION:
: APPLICANT: Conklin, Darrell C.
: APPLICANT: Lottion-Day, Catherine E.
: APPLICANT: Lok, Si
: APPLICANT: Jaspers, Stephen R.
: TITLE OF INVENTION: INSULIN HOMOLOG
: NUMBER OF SEQUENCES: 17
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Zymogenetics, Inc.
: STREET: 1201 Eastlake Avenue East
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98102
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSEQ for Windows Version 2.0

```

1 CURRENT APPLICATION DATA:  
2 APPLICATION NUMBER: US/08/950,720A  
3 FILING DATE:  
4 CLASSIFICATION: 435  
5 PRIOR APPLICATION DATA:  
6 APPLICATION NUMBER:  
7 FILING DATE:  
8 ATTORNEY/AGENT INFORMATION:  
9 NAME: Sawislak, Deborah A  
10 REGISTRATION NUMBER: 37,438  
11 REFERENCE/DOCKET NUMBER: 96-09  
12 TELECOMMUNICATION INFORMATION:  
13 TELEPHONE: 206-442-6672  
14 TELEFAX: 206-442-6678  
15 TELEX:  
16 INFORMATION FOR SEQ ID NO: 11:  
17 SEQUENCE CHARACTERISTICS:  
18 LENGTH: 110 amino acids  
19 TYPE: amino acid  
20 STRANDEDNESS: single  
21 TOPOLOGY: linear  
22 MOLECULE TYPE: protein  
23 US-08-950-720A 11

Query Match 100.0% Score 464; DB 3; Length 110;  
Best Local Similarity 100.0% Pred No. 2,1e-47;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAFDLQVGQVGLGGPGAGSLQPLALEG 60  
DB 25 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAFDLQVGQVGLGGPGAGSLQPLALEG 84  
QY 61 SLQKRGIVEOCCTSGCSLYOLENYCN 86  
DB 85 SLQKRGIVEOCCTSGCSLYOLENYCN 110

RESULT 9  
US-08-589-028-2  
Sequence 2, Application US/08589028  
Patent No. 6087129  
GENERAL INFORMATION:  
APPLICANT: Newgard, Christopher B.  
APPLICANT: Halban, Philippe A.  
APPLICANT: No. 6087129minington, Karl D.  
APPLICANT: Clark, Samuel A.  
APPLICANT: Thiippen, Anice E.  
APPLICANT: Quade, Christian  
APPLICANT: Kruse, Fred  
TITLE OF INVENTION: Recombinant Expression of Proteins From  
TITLE OF INVENTION: SECRETORY CELL LINES  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P. O. Box 4433  
CITY: Houston  
STATE: TX  
COUNTRY: USA  
ZIP: 77210-4433  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/589,028  
FILING DATE: Concurrently Herewith  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Highlander, Steven L.  
REGISTRATION NUMBER: 47,642  
REFERENCE/DOCKET NUMBER: 47,642  
TELECOMMUNICATION INFORMATION:

1 TELEPHONE: (512) 418-3000  
2 TELEFAX: (512) 474-7577  
3 INFORMATION FOR SEQ ID NO: 2:  
4 SEQUENCE CHARACTERISTICS:  
5 LENGTH: 110 amino acids  
6 TYPE: amino acid  
7 STRANDEDNESS:  
8 TOPOLOGY: linear  
9 US-08-589-028-2  
Query Match 100.0% Score 463; DB 3; Length 110;  
Best Local Similarity 100.0% Pred No. 2,1e-47;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAFDLQVGQVGLGGPGAGSLQPLALEG 60  
DB 25 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAFDLQVGQVGLGGPGAGSLQPLALEG 84  
QY 61 SLQKRGIVEOCCTSGCSLYOLENYCN 86  
DB 85 SLQKRGIVEOCCTSGCSLYOLENYCN 110

RESULT 10  
US-08-784-582-2  
Sequence 2, Application US/08784582  
Patent No. 6110707  
GENERAL INFORMATION:  
APPLICANT: Newgard, Christopher B.  
APPLICANT: Halban, Philippe A.  
APPLICANT: No. 6110707minington, Karl D.  
APPLICANT: Clark, Samuel A.  
APPLICANT: Thiippen, Anice E.  
APPLICANT: Quade, Christian  
APPLICANT: Kruse, Fred  
APPLICANT: McGarry, Dennis  
TITLE OF INVENTION: RECOMBINANT EXPRESSION OF PROTEINS FROM  
TITLE OF INVENTION: SECRETORY CELL LINES  
NUMBER OF SEQUENCES: 79  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P. O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/784,582  
FILING DATE: Concurrently Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/026,427  
FILING DATE: 15-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/589,028  
FILING DATE: 19-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Highlander, Steven L.  
REGISTRATION NUMBER: 37,642  
REFERENCE/DOCKET NUMBER: UTSD:514  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 512/418-3000  
TELEFAX: 512/474-7577  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 110 amino acids  
TYPE: amino acid  
STRANDEDNESS:

```

QY 61 SLOKRGIVEQCCTSCSYQLENYCN 86
      |||||||
DB 85 SLOKRGIVEQCCTSCSYQLENYCN 110

RESULT 12
US-08-472-701-2
: Sequence 2, Application US/08472701
: Patent No. 6509165
: GENERAL INFORMATION:
: APPLICANT: Griffin, Ann C.
: APPLICANT: Hickey, William F.
: TITLE OF INVENTION: Detection and Treatment Methods for
: TYPE OF INVENTION: Type I Diabetes
: NUMBER OF SEQUENCES: 23
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: LAHIVE & COCKFIELD
: STREET: 60 State Street, suite 510
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: USA
: ZINC: 04109 1876
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC Compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: ASCII Text
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/472,701
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/272,220
: FILING DATE: 08-JULY-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: DeConti, Giulio A., Jr.
: REGISTRATION NUMBER: 31,531
: REFERENCE/DOCKET NUMBER: DC1-092DV
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617)227-7400
: TELEFAX: (617)227-5941
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 110 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-472-701-2

Query Match 100.0%; Score 463; DB 4; Length 110;
Best Local Similarity 100.0%; Pred. No. 2,le-47;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 1 FVNHQHGSHLVEALVYCGERGFFYTKTKREAEDLVGVGVGGPGAGSLQPLALEG 60
      |||||||
DB 25 FVNHQHGSHLVEALVYCGERGFFYTKTKREAEDLVGVGVGGPGAGSLQPLALEG 84

QY 61 SLOKRGIVEQCCTSCSYQLENYCN 86
      |||||||
DB 85 SLOKRGIVEQCCTSCSYQLENYCN 110

RESULT 13
US-09-185-852-2
: Sequence 2, Application US/09185852
: Patent No. 6537806
: GENERAL INFORMATION:
: APPLICANT: Osborne, William R.A.
: APPLICANT: Ramesh, Nagarajan
: TITLE OF INVENTION: Compositions and Methods for Treating Diabetes
: FILE REFERENCE: P-UW 3264
: CURRENT APPLICATION NUMBER: US/09/185,852

```

```

: CURRENT FILING DATE: 1998-11-04
: EARLIER APPLICATION NUMBER: 60/087,660
: EARLIER FILING DATE: 1998-06-02
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 2
: LENGTH: 110
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-185-852-2

Query Match 100.08; Score 463; DB 4; Length 110;
Best Local Similarity 100.08; Pred. No. 2,1e-47;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
      |||||
Db 25 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 84
      |||||
QY 61 SLOKRGIVEQCCTISCSLYOLENYCN 86
      |||||
Db 85 SLOKRGIVEQCCTISCSLYOLENYCN 110
      |||||

RESULT 15
US-09-280-030-63
: Sequence 63; Application US/09280030A
: Patent No. 6506595
: GENERAL INFORMATION:
: APPLICANT: Sato, Seiji
: APPLICANT: Higashikuni, Naohiko
: APPLICANT: Kudo, Toshiyuki
: APPLICANT: Kondo, Masaaki
: TITLE OF INVENTION: DNAS ENCODING NEW FUSION PROTEINS AND PROCESSES FOR
: TITLE OF INVENTION: PREPARING USEFUL POLYPEPTIDES THROUGH EXPRESSION OF THE
: TITLE OF INVENTION: DNAS
: FILING REFERENCE: 382,1826
: CURRENT APPLICATION NUMBER: US/09/280,030 A
: EARLIER FILING DATE: 1999-04-26
: EARLIER APPLICATION NUMBER: 3916-87339/1998
: NUMBER OF SEQ ID NOS: 65
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 63
: LENGTH: 117
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Designated as
: OTHER INFORMATION: an amino acid sequence of
: OTHER INFORMATION: MW:sp:MW:mp:10 Met-Proinsulin
US-09-280-030-63

Query Match 100.08; Score 463; DB 4; Length 117;
Best Local Similarity 100.08; Pred. No. 2,2e-47;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
      |||||
Db 32 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 91
      |||||
QY 61 SLOKRGIVEQCCTISCSLYOLENYCN 86
      |||||
Db 92 SLOKRGIVEQCCTISCSLYOLENYCN 117
      |||||

Search completed: September 15, 2003, 12:05:32
Job time : 20.7276 secs

Query Match 100.08; Score 463; DB 5; Length 110;
Best Local Similarity 100.08; Pred. No. 2,1e-47;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
      |||||

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein : protein search, using sw model

Run on: September 15, 2003, 12:03:35 : Search time 36.0645 seconds  
(without alignments)  
347.945 Million cell updates/sec

Title: US-09-423-100-4

Perfect score: 463

Sequence: 1 FVNHLCGSHLVYALVLCG.....IVEQCCTSCSLYLENYCN 86

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 541936 seqs, 145912426 residues

Total number of hits satisfying chosen parameters: 541936

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/1/pubpa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpa/US09A\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpa/US10H\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	463	100.0	86	10	US-09-878-380-1
2	463	100.0	86	11	US-09-858-935B-4
3	463	100.0	86	14	US-10-028-115-2
4	463	100.0	86	14	US-10-054-873-4
5	463	100.0	96	10	US-09-947-663-4
6	463	100.0	110	9	US-09-295-8-125
7	463	100.0	110	9	US-09-815-229-3
8	463	100.0	110	10	US-09-804-409A-9
9	463	100.0	110	12	US-09-969-748C-6
10	463	100.0	110	15	US-10-038-686-1
11	463	100.0	110	15	US-10-328-813-2
12	463	100.0	117	9	US-09-280-030-63
13	463	100.0	130	9	US-09-280-030-62
14	457	98.7	96	10	US-09-947-563-5
15	306	66.1	166	9	US-09-925-297-805

16	300	64.8	56	9	US-09-815-229-10	Sequence 10, Appl
17	285	61.6	54	9	US-09-815-229-13	Sequence 13, Appl
18	267	57.7	52	14	US-10-054-873-5	Sequence 5, Appl
19	267	57.7	107	14	US-10-054-873-6	Sequence 6, Appl
20	267	57.7	150	14	US-10-054-873-7	Sequence 7, Appl
21	261.5	56.5	51	11	US-09-858-935B-5	Sequence 5, Appl
22	261.5	56.5	51	14	US-10-028-410-3	Sequence 3, Appl
23	258.5	55.8	124	10	US-09-894-711-18	Sequence 18, Appl
24	258.5	55.8	138	10	US-09-861-687-39	Sequence 19, Appl
25	256	55.3	50	14	US-10-066-009A-3	Sequence 3, Appl
26	250.5	54.1	124	9	US-09-736-611-12	Sequence 12, Appl
27	250.5	54.1	125	9	US-09-736-611-10	Sequence 10, Appl
28	250.5	54.1	147	9	US-09-736-611-8	Sequence 8, Appl
29	249.5	53.9	124	9	US-09-740-359-12	Sequence 12, Appl
30	249.5	53.9	124	10	US-09-894-711-12	Sequence 12, Appl
31	249.5	53.9	125	9	US-09-740-359-10	Sequence 10, Appl
32	249.5	53.9	125	10	US-09-894-711-10	Sequence 10, Appl
33	249.5	53.9	147	9	US-09-740-359-7	Sequence 7, Appl
34	248	53.6	144	9	US-09-746-611-6	Sequence 6, Appl
35	248	53.6	144	9	US-09-740-359-5	Sequence 5, Appl
36	248	53.6	146	10	US-09-894-711-5	Sequence 5, Appl
37	227.5	49.1	144	10	US-09-894-711-7	Sequence 7, Appl
38	209	45.3	46	9	US-09-205-658-134	Sequence 132, App
39	197	42.5	46	9	US-09-205-658-133	Sequence 133, App
40	184.5	39.8	50	9	US-09-205-658-115	Sequence 115, App
41	180	38.9	46	9	US-09-205-658-135	Sequence 135, App
42	178	38.4	42	9	US-09-815-229-14	Sequence 14, Appl
43	178	38.4	32	10	US-09-947-563-6	Sequence 6, Appl
44	175	37.8	46	9	US-09-205-658-134	Sequence 134, App
45	174	37.6	35	9	US-09-815-229-4	Sequence 4, Appl

#### ALIGNMENTS

RESULT 1  
US-09-878-380-1  
Sequence 1: Application US/09878380  
Patent No: US20020160435A1  
GENERAL INFORMATION:  
APPLICANT: Fujirebio Inc.  
APPLICANT: KITAJIMA, Sachiko  
APPLICANT: KURANO, Yoshihiro  
APPLICANT: NAKATSUBO, Kaoru  
APPLICANT: NISHIZONO, Isao  
TITLE OF INVENTION: Immunoassay For Measuring Human C-Peptide and Kit Therefor  
FILE REFERENCE: 0760-0291P  
CURRENT APPLICATION NUMBER: US/09/878,380  
PRIOR FILING DATE: 2001-06-12  
PRIOR APPLICATION NUMBER: JP 2000-174691  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1  
LENGTH: 86  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-878-380-1

Query Match 100.0% Score 463; DB 10: Length 86;  
Best Local Similarity 100.0% Pred. No. 9.7e 45;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNHLCGSHLVYALVLCGGEFFYTPKTRAEADLVGVGVGGFGAGSLQIALEG 60  
|||||  
Db 1 FVNHLCGSHLVYALVLCGGEFFYTPKTRAEADLVGVGVGGFGAGSLQIALEG 60  
Qy 61 SLOKRGIVEQCCTSCSLYLENYCN 86  
|||||  
Db 61 SLOKRGIVEQCCTSCSLYLENYCN 86

RESULT 2

```

US-09-858-945B-4
: Sequence 4, Application US/09858935B
: Publication No. US20030069177A1
: GENERAL INFORMATION:
: APPLICANT: Dubaqui, Yves
: APPLICANT: Filvaroff, Ellen
: TITLE OF INVENTION: METHOD FOR TREATING CARTILAGE DISORDERS
: FILE REFERENCE: P1794R1
: CURRENT APPLICATION NUMBER: US/09/858,935B
: PRIOR FILING DATE: 2002-07-02
: PRIOR APPLICATION NUMBER: US 60/248,985
: PRIOR FILING DATE: 2000-11-15
: PRIOR APPLICATION NUMBER: US 60/204,450
: PRIOR FILING DATE: 2000-05-16
: NUMBER OF SEQ ID NOS: 153
: SEQ ID NO 4
: LENGTH: 86
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-858-945B-4

Query Match 100.0% Score 454 DB 14 Length 86
Best Local Similarity 100.0% Pred. No. 9,7e-45
Matches 86 Conservative 0 Mismatches 0 Indels 0 Gaps 0

QY 1 FVNHLGSHLVEALYLVCGERGFFYTPKTRKAEADLVGVGVGQPGAGSLQPLALEG 60
DB 1 FVNHLGSHLVEALYLVCGERGFFYTPKTRKAEADLVGVGVGQPGAGSLQPLALEG 60

QY 61 SLKRGIVGQCCTSGISLYOLENYCN 86
DB 61 SLKRGIVGQCCTSGISLYOLENYCN 86

RESULT 4
US-10-028-410-2
: Sequence 2, Application US/10028410
: Publication No. US20020160955A1
: GENERAL INFORMATION:
: APPLICANT: Dubaqui, Yves
: APPLICANT: Lowman, Henry
: TITLE OF INVENTION: PROTEIN VARIANTS
: FILE REFERENCE: P1712R1-1
: CURRENT APPLICATION NUMBER: US/10/028,410
: CURRENT FILING DATE: 2001-12-19
: PRIOR APPLICATION NUMBER: US/09/473,524
: PRIOR FILING DATE: 2000-01-05
: NUMBER OF SEQ ID NOS: 6
: SEQ ID NO 2
: LENGTH: 86
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-10-028-410-2

Query Match 100.0% Score 454 DB 14 Length 86
Best Local Similarity 100.0% Pred. No. 9,7e-45
Matches 86 Conservative 0 Mismatches 0 Indels 0 Gaps 0

QY 1 FVNHLGSHLVEALYLVCGERGFFYTPKTRKAEADLVGVGVGQPGAGSLQPLALEG 60
DB 1 FVNHLGSHLVEALYLVCGERGFFYTPKTRKAEADLVGVGVGQPGAGSLQPLALEG 60

QY 61 SLKRGIVGQCCTSGISLYOLENYCN 86
DB 61 SLKRGIVGQCCTSGISLYOLENYCN 86

RESULT 5
US-25-947-563-4
: Sequence 4, Application US/09947563
: Patent No. US20020156254A1
: GENERAL INFORMATION:
: APPLICANT: Rubroeder, Franz-Josef
: APPLICANT: Keller, Reinhold
: TITLE OF INVENTION: Improved process for obtaining
: insulin precursors having correctly bonded cystine bridge
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Finnegan, Henderson, Parrabow, Garrett &
: Dunner
: STREET: 1300 I Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-3315
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION NUMBER: US/09/947,563  
FILING DATE: 07-Sep-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/134,836  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Leslie McDonnell  
REGISTRATION NUMBER: 34,872  
REFERENCE/DOCKET NUMBER: 02481.16.00 00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 96 amino acids  
TYPE: amino acid  
SUBSTANCE: Simvastatin  
LOCALITY: Linear  
MOLECULAR TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Escherichia coli  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..96  
SEQUENCE DESCRIPTION: SEQ ID NO: 4  
US-09-947-563-4

Query Match 100.0%; Score 463; DB 10; Length 96;  
Best Local Similarity 100.0%; Pred. No. 1.1e-44;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPGAGSLQPLALEG 60  
|||||  
DB 11 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPGAGSLQPLALEG 70  
|||||

QY 61 SLOKRGIVEQCCTSCISLYQLENYCN 86  
|||||  
DB 71 SLOKRGIVEQCCTSCISLYQLENYCN 96  
|||||

RESULT 6  
US-09-205-658-125  
Sequence 125, Application US/09205658  
Patent No. US20010029617A1  
GENERAL INFORMATION:  
APPLICANT: Ruvkun, Gary  
APPLICANT: Cqq, Scott  
TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR  
TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS  
FILE REFERENCE: 00786/351004  
CURRENT APPLICATION NUMBER: US/09/205,658  
CURRENT FILING DATE: 1998-12-03  
EARLIER APPLICATION NUMBER: 08/857,076  
EARLIER FILING DATE: 1997-05-15  
EARLIER APPLICATION NUMBER: 08/888,514  
EARLIER FILING DATE: 1997-07-07  
EARLIER APPLICATION NUMBER: US98/10080  
EARLIER FILING DATE: 1998-05-15  
NUMBER OF SEQ ID NOS: 328  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 125  
LENGTH: 110  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-205-658-125

Query Match 100.0%; Score 463; DB 9; Length 110;  
Best Local Similarity 100.0%; Pred. No. 1.3e-44;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPGAGSLQPLALEG 60  
|||||  
DB 25 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPGAGSLQPLALEG 84  
|||||

QY 61 SLOKRGIVEQCCTSCISLYQLENYCN 86  
|||||  
DB 85 SLOKRGIVEQCCTSCISLYQLENYCN 110  
|||||

RESULT 7  
US-09-815-229-3  
Sequence 3, Application US/09815229  
Patent No. US20020058614A1  
GENERAL INFORMATION:  
APPLICANT: Filvaroff, Ellen H.  
APPLICANT: Okumu, Franklin W.  
TITLE OF INVENTION: USE OF INSULIN FOR THE TREATMENT OF CARTILAGENOUS DISORDERS  
FILE REFERENCE: P1786RIUS  
CURRENT APPLICATION NUMBER: US/09/815,229  
CURRENT FILING DATE: 2001-03-22  
PRIOR APPLICATION NUMBER: US 60/192,173  
PRIOR FILING DATE: 2000-03-24  
NUMBER OF SEQ ID NOS: 17  
SEQ ID NO 3  
LENGTH: 110  
TYPE: PRT  
ORGANISM: Homo sap ius  
US-09-815-229-3

Query Match 100.0%; Score 463; DB 9; Length 110;  
Best Local Similarity 100.0%; Pred. No. 1.3e-44;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPGAGSLQPLALEG 60  
|||||  
DB 25 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPGAGSLQPLALEG 84  
|||||

QY 61 SLOKRGIVEQCCTSCISLYQLENYCN 86  
|||||  
DB 85 SLOKRGIVEQCCTSCISLYQLENYCN 110  
|||||

RESULT 8  
US-09-804-409A-9  
Sequence 9, Application US/09804409A  
Patent No. US20020155100A1  
GENERAL INFORMATION:  
APPLICANT: KIERFER, TIMOTHY J.  
APPLICANT: CHEUNG, ANTHONY T.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATED PROTEIN  
TITLE OF INVENTION: EXPRESSION IN GUT  
FILE REFERENCE: 029996/027 8721  
CURRENT APPLICATION NUMBER: US/09/804,409A  
CURRENT FILING DATE: 2001-03-12  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 9  
LENGTH: 110  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-804-409A-9

Query Match 100.0%; Score 463; DB 10; Length 110;  
Best Local Similarity 100.0%; Pred. No. 1.3e-44;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPGAGSLQPLALEG 60  
|||||  
DB 25 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPGAGSLQPLALEG 84  
|||||

QY 61 SLOKRGIVEQCCTSCISLYQLENYCN 86  
|||||

Db 85 SLOKRGIVECCCTSCISLYOLENYCN 110

## RESULT 9

US-09-969-748C-6  
Sequence 6, Application US/09969748C  
Publication No. US20030161809A1

## GENERAL INFORMATION:

APPLICANT: ARIZERE PHARMACEUTICALS, INC.  
APPLICANT: HOUSTON, Lou, L.  
APPLICANT: SHERIDAN, Philip, J.  
APPLICANT: HAWLEY, Stephen  
APPLICANT: GLYNN, Jacqueline, M.  
APPLICANT: CHAPIN, Steven  
APPLICANT: BASU, Amresh

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TRANSPORT OF BIOLOGICALLY ACTIVE AGENTS ACROSS CELLULAR BARRIERS

FILE REFERENCE: 057220-0303

CURRENT APPLICATION NUMBER: US 60/969,748C

PRIOR FILING DATE: 2002-12-16

PRIOR APPLICATION NUMBER: US 60/267,666

PRIOR FILING DATE: 2001-02-24

PRIOR APPLICATION NUMBER: US 60/245,819

PRIOR FILING DATE: 2000-11-14

PRIOR APPLICATION NUMBER: US 60/248,478

PRIOR FILING DATE: 2000-11-13

PRIOR APPLICATION NUMBER: US 60/237,929

PRIOR FILING DATE: 2000-10-02

NUMBER OF SEQ ID NOS: 115

SOFTWARE: PatentIn version 4.0

SEQ ID NO: 6

LENGTH: 110

TYPE: PRT

ORGANISM: Homo sapiens

US-09-969-748C-6

Query Match 100.0%; Score 463; DB 12; Length 110;  
Best Local Similarity 100.0%; Pred. No. 1.3e-44;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FVNOHLCGSHLVEALYVCGERGFFYTPKTRREAEDLVQGVGLGGPGAGSLOPLALEG 60

Db 25 FVNOHLCGSHLVEALYVCGERGFFYTPKTRREAEDLVQGVGLGGPGAGSLOPLALEG 84

OY 61 SLOKRGIVECCCTSCISLYOLENYCN 86

Db 85 SLOKRGIVECCCTSCISLYOLENYCN 110

## RESULT 10

US-10-038-686-1  
Sequence 1, Application US/10038686  
Publication No. US20030045467A1

## GENERAL INFORMATION:

APPLICANT: Orban, Tibamir  
TITLE OF INVENTION: AUTOANTIGEN VACCINE

FILE REFERENCE: 10276-067001

CURRENT APPLICATION NUMBER: US/10/038,686

PRIOR FILING DATE: 2002-05-31

PRIOR APPLICATION NUMBER: 60/266,068

PRIOR FILING DATE: 2001-01-05

NUMBER OF SEQ ID NOS: 6

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO: 1

LENGTH: 110

TYPE: PRT

ORGANISM: Homo sapiens

US-10-038-686-1

Query Match 100.0%; Score 463; DB 15; Length 110;  
Best Local Similarity 100.0%; Pred. No. 1.3e-44;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FVNOHLCGSHLVEALYVCGERGFFYTPKTRREAEDLVQGVGLGGPGAGSLOPLALEG 60  
Db 25 FVNOHLCGSHLVEALYVCGERGFFYTPKTRREAEDLVQGVGLGGPGAGSLOPLALEG 84

OY 61 SLOKRGIVECCCTSCISLYOLENYCN 86

Db 85 SLOKRGIVECCCTSCISLYOLENYCN 110

## RESULT 11

US-10-328-813-2

Sequence 2, Application US/10328813

Publication No. US20030113305A1

## GENERAL INFORMATION:

APPLICANT: Osbourne, William R.A.

APPLICANT: Ramesh, Nagarajan

TITLE OF INVENTION: Compositions and Methods for Treating Diabetes

FILE REFERENCE: P-W 3264

CURRENT APPLICATION NUMBER: US/10/328,813

PRIOR FILING DATE: 2002-12-25

PRIOR APPLICATION NUMBER: US/09/785,452

PRIOR FILING DATE: 1998-11-04

PRIOR APPLICATION NUMBER: 60/087,666

NUMBER OF SEQ ID NOS: 11

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO: 2

LENGTH: 110

TYPE: PRT

ORGANISM: Homo sapiens

US-10-328-813-2

Query Match 100.0%; Score 454; DB 15; Length 110;  
Best Local Similarity 100.0%; Pred. No. 1.3e-44;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FVNOHLCGSHLVEALYVCGERGFFYTPKTRREAEDLVQGVGLGGPGAGSLOPLALEG 60

Db 25 FVNOHLCGSHLVEALYVCGERGFFYTPKTRREAEDLVQGVGLGGPGAGSLOPLALEG 84

OY 61 SLOKRGIVECCCTSCISLYOLENYCN 86

Db 85 SLOKRGIVECCCTSCISLYOLENYCN 110

## RESULT 12

US-09-280-030-63

Sequence 63, Application US/09280030A

Patent No. US20010021515A1

## GENERAL INFORMATION:

APPLICANT: Sato, Seiji

APPLICANT: Higashikuni, Naohiko

APPLICANT: Kudo, Toshiyuki

APPLICANT: Kondo, Masaaki

TITLE OF INVENTION: DNAs ENCODING NEW FUSION PROTEINS AND PROCESSES FOR

TITLE OF INVENTION: PREPARING USEFUL POLYPEPTIDES THROUGH EXPRESSION OF THE

FILE REFERENCE: 382 1026

CURRENT APPLICATION NUMBER: US/09/280,030A

CURRENT FILING DATE: 1999-03-26

EARLIER APPLICATION NUMBER: JP10-87339/1998

EARLIER FILING DATE: 1998-03-31

NUMBER OF SEQ ID NOS: 66

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO: 63

LENGTH: 117

TYPE: PRT

ORGANISM: Artificial: Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Designated is

OTHER INFORMATION: an amino acid sequence of

OTHER INFORMATION: MW:sp-MWmp10-Met-Proinsulin

US-09-280-030-63

```

Query Match          100.0%; Score 463; DB 9; Length 117;
Best Local Similarity 100.0%; Pred. No. 1.4e-44;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRKREAEADLVQGVLEGGPGAGSLQPIALEG 60
   |||||||
Db 32 FVNHLCGSHLVEALYLVCGERGFFYTPKTRKREAEADLVQGVLEGGPGAGSLQPIALEG 91
   |||||||

QY 61 SLOKRGIVEQCCTSCSLYLENYCN 86
   |||||||
Db 92 SLOKRGIVEQCCTSCSLYLENYCN 117
   |||||||

RESULT 13
US-09-280-030-62
: Sequence 62, Application US/09280030A
: Patent No. US20016021515A1
: GENERAL INFORMATION:
: APPLICANT: Saito, Seiji
: APPLICANT: Blazshikuni, Naoki-ko
: APPLICANT: Kudo, Toshiyuki
: APPLICANT: Kado, Masaki
: TITLE OF INVENTION: DNAS ENCODING NEW HES-ON PROTEINS AND PROCESSES FOR
: TITLE OF INVENTION: PREPARING USEFUL POLYPEPTIDES THROUGH EXPRESSION OF THE
: FILE REFERENCE: 382.1026
: CURRENT APPLICATION NUMBER: US/09/280,0
: CURRENT FILING DATE: 1999-03-26
: EARLIER APPLICATION NUMBER: JP10-87319/
: EARLIER FILING DATE: 1998-03-11
: NUMBER OF SEQ ID NOS: 66
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 62
: LENGTH: 130
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Designated is
: OTHER INFORMATION: an amino acid sequence of
: OTHER INFORMATION: MWPsp-MWmp10-(His)6-Linker-Met-Proinsulin
US-09-280-030-62

Query Match          100.0%; Score 463; DB 9; Length 130;
Best Local Similarity 100.0%; Pred. No. 1.6e-44;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRKREAEADLVQGVLEGGPGAGSLQPIALEG 60
   |||||||
Db 45 FVNHLCGSHLVEALYLVCGERGFFYTPKTRKREAEADLVQGVLEGGPGAGSLQPIALEG 104
   |||||||

QY 61 SLOKRGIVEQCCTSCSLYLENYCN 86
   |||||||
Db 105 SLOKRGIVEQCCTSCSLYLENYCN 130
   |||||||

RESULT 14
US-09-947-563-5
: Sequence 5, Application US/09947563
: Patent No. US2002156234A1
: GENERAL INFORMATION:
: APPLICANT: Rubroder, Franz-Josef
: APPLICANT: Keller, Reinhold
: TITLE OF INVENTION: Improved process for obtaining
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
: Dunner
: STREET: 1300 I Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA

```

```

ZIP: 20005-3715
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/947,563
FILING DATE: 07-Sep-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/134,836
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Leslie McDonnell
REGISTRATION NUMBER: 34,872
REFERENCE/DOCKET NUMBER: 62481.1600.00306
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
E-MAIL FOR SEQ ID NOS: 51
SEQUENCE CHARACTERISTICS:
LENGTH: 94 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: Protein
LOCATION: 1..96
SEQUENCE DESCRIPTION: SEQ ID NO: 51
US-09-947-563-5

Query Match          94.7%; Score 457; DB 10; Length 96;
Best Local Similarity 100.0%; Pred. No. 5.2e-44;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRKREAEADLVQGVLEGGPGAGSLQPIALEG 60
   |||||||
Db 11 FVNHLCGSHLVEALYLVCGERGFFYTPKTRKREAEADLVQGVLEGGPGAGSLQPIALEG 70
   |||||||

QY 61 SLOKRGIVEQCCTSCSLYLENYCN 85
   |||||||
Db 71 SLOKRGIVEQCCTSCSLYLENYCN 95
   |||||||

RESULT 15
US-09-925-297-805
: Sequence 805, Application US/09925297
: Patent No. US20020081659A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
: FILE REFERENCE: PA105
: CURRENT APPLICATION NUMBER: US/09/925,297
: CURRENT FILING DATE: 2001-08-10
: PRIOR APPLICATION NUMBER: PCT/US00/05989
: PRIOR FILING DATE: 2000-03-08
: PRIOR APPLICATION NUMBER: 60/124,270
: PRIOR FILING DATE: 1999-03-12
: NUMBER OF SEQ ID NOS: 928
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 805
: LENGTH: 166
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: SITE
: LOCATION: (84)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE

```

```
; LOCATION: (92)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (105)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (124)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (132)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (144)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (145)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (165)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
US-09-423-297-805
```

```
Query Match: 56.1%; Score: 300; 18 hits; Length: 176.
Best Local Similarity: 96.7%; Pred. No. 5; 2e-27;
Matches: 58; Conservative: 0; Mismatch: 3; 2; Indels: 0; Gaps: 0;
Oy 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRPEADLQVGQVELGGGPGAGSIQPLEAES 60
Db 42 FVNHLCGSHLVEALYLVCGERGFFYTPKTRPEADLQVGQVELGGGPGAGSIQPLEAES 101
```

Search completed: September 15, 2003, 12:24:35  
Job time : 46.0645 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

GM protein - protein search, using sw model

Run on: September 15, 2003, 11:56:45 / Search time 14.1792 seconds  
(without alignments)  
583.284 Million cell updates/sec

Title: US-09-423-100-4

Perfect score: 453

Sequence: 1 FVNOHLCGSHAEVLYVGG.....IVPGQCTSGISLYGLENYIN 86

Scoring table: BL2SUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 95168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB Seq Length: 0

Maximum DB Seq Length: 200000000

Post-processing: Minimum Match: 0.5

Maximum Match: 100%

Listing first 45 summaries

Database : PIR76:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	463	100.0	110	1 IPHG	insulin precursor
2	461	100.0	110	2 A42179	insulin precursor
3	456	98.5	110	2 B42179	insulin precursor
4	456	98.5	110	2 J00178	insulin precursor
5	424	91.6	110	3 INK6	insulin precursor
6	417	90.1	110	1 IPG5	insulin precursor
7	414	85.1	86	3 IPH0	insulin precursor
8	404	85.1	110	1 IPK72	insulin precursor
9	404	85.1	110	1 INNS2	insulin 2 precursor
10	392	84.7	108	2 A39883	insulin precursor
11	392	84.7	110	2 I48166	insulin precursor
12	385	83.2	110	1 IPK1	insulin 1 precursor
13	383	82.7	84	1 IPG3	insulin precursor
14	366.5	79.2	105	1 IPB9	insulin precursor
15	346	79.0	108	1 INNS1	insulin precursor
16	334.5	72.2	108	2 S02278	insulin precursor
17	320.5	69.2	77	1 INSH	insulin precursor
18	314	67.8	110	1 IPV3	insulin precursor
19	277.5	59.9	109	1 IPK10	insulin precursor
20	276.5	59.7	103	2 I51221	insulin precursor
21	265.5	57.3	106	1 IPXL2	insulin II precursor
22	265.5	57.3	107	1 IPCH	insulin precursor
23	262.5	56.7	106	1 IPXL1	insulin 1 precursor
24	256.5	55.4	51	1 INNHP	insulin - sperm wh
25	256.5	55.4	51	1 INNH	insulin - finback
26	256.5	55.4	51	1 INEL	insulin - elephant
27	256.5	55.4	81	1 IPK	insulin precursor
28	256	55.3	96	2 PC7082	epidermal growth f
29	254.5	55.0	51	1 INHY	insulin - hamster

insulin - Egyptian  
insulin precursor  
insulin - sei whal  
insulin - goat  
insulin - Arabian  
insulin - cat  
insulin - Common s  
insulin - North Am  
insulin - Chinchip  
insulin - goose  
insulin - crested  
insulin - turkey ( ostrich  
insulin - black be  
insulin - slider t  
insulin II - North

## ALIGNMENTS

### RESULT 1

IPH1

insulin precursor [artificial] - human

N. Alternans: Babes: preproinsulin

C. Species: Homo sapiens (man)

C. Date: 23-Oct-1981 #sequence\_revision 23 Oct 1981 #text\_change 08-Dec-2000

C. Accession: A93222; A94253; A94216; A94251; A93144; A92075; A9186; 158114; A01579  
R. Bell, G. L. Pictet, R. L. Rutter, W. J. Cordell, B. Tischer, E. Goodman, H. M.

Nature 284, 26-32, 1980

A. Title: Sequence of the human insulin gene.

A. Reference number: A93222; M01180126725; PMID:6244748

A. Accession: A93222

A. Molecule type: DNA

A. Residues: 1-110 <REL>

A. Cross-references: GR:J00265; NID:q186429; PIDN:AAA59172.1; PID:q336828

R. Ullrich, A. Dull, F. J. Gray, A. Brosius, J. Sures, I.

Science 209, 612-615, 1980

A. Title: Genetic variation in the human insulin gene.

A. Reference number: A94253; M01080236313; PMID:6248962

A. Accession: A94253

A. Molecule type: DNA

A. Residues: 1-110 <REL>

A. Cross-references: GR:J00265; NID:q186429; PIDN:AAA59172.1; PID:q336828

R. Bell, G. L. Pictet, R. L. Rutter, W. J. Cordell, B. Tischer, E. Goodman, H. M.; Rutter, W. J.

Nature 282, 525-527, 1979

A. Title: Nucleotide sequence of a cDNA clone encoding human preproinsulin.

A. Reference number: A93216; M01080054779; PMID:503324

A. Accession: A93216

A. Molecule type: mRNA

A. Residues: 1-110 <REL>

A. Cross-references: GR:J00265; NID:q186429; PIDN:AAA59172.1; PID:q336828

R. Sures, I. Ullrich, D. V. Gray, A. Ullrich, A.

Science 208, 57-59, 1980

A. Title: Nucleotide sequence of human preproinsulin complementary DNA.

A. Reference number: A94251; M01080147417; PMID:6927840

A. Accession: A94251

A. Molecule type: mRNA

A. Residues: 1-110 <REL>

A. Cross-references: GR:J00265; NID:q186429; PIDN:AAA59172.1; PID:q336828

R. Nicol, D. S. H. W. Smith, L. F.

Nature 187, 483-485, 1960

A. Title: Amino acid sequence of human insulin.

A. Reference number: A93144

A. Accession: A93144

A. Molecule type: protein

A. Residues: 25-54; 90-110 <NIC>

R. Oyer, P. E. Cho, S. Peterson, J. D. Steiner, D. F.

J. Biol. Chem. 246, 1375-1386, 1971

A. Title: Studies on human proinsulin. Isolation and amino acid sequence of the huma

A. Reference number: A92075; M0107116410; PMID:5101771

A. Accession: A92075

A. Molecule type: protein

A:Residues: 57-87 <OEV>  
 R:Ko, A.; Smyth, D.G.; Markussen, J.; Sundby, F.  
 Eur. J. Biochem. 20, 190-199, 1971  
 A:Title: Amino acid sequence of the C-peptide of human proinsulin.  
 A:Accession: A91186; MUID:71257722; PMID:5560404  
 A:Reference number: A91186; MUID:71257722; PMID:5560404  
 A:Molecule type: protein  
 A:Residues: 57-87 <KOA>  
 R:Lucassen, A.M.; Julier, C.; Beressi, J.P.; Moitard, C.; Froguet, P.; Lathrop, M.; Bell  
 Nature Genet. 4, 305-310, 1993  
 A:Title: Susceptibility to insulin dependent diabetes mellitus maps to a 4.1 kb segment  
 A:Reference number: 158114; MUID:93364428; PMID:8358440  
 A:Accession: 158114  
 A:Status: preliminary; translated from GR/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-59, 63-110 <RFS>  
 A:Cross-references: GR:115440; NID:q307071; PIDN:AA59179.1; PID:q307072  
 R:Sieber, P.; Kamber, B.; Hartmann, A.; Joehli, A.; Rinkler, B.; Ritte, W.  
 Helv. Chim. Acta 57, 2617-2621, 1974  
 A:Title: Totalsynthese von Humaninsulin unter azobischer Bildung der Disulfidbindung.  
 A:Reference number: A91636; MUID:75077273; PMID:444424  
 A:Contents: annotation; synthesis  
 A:Note: disulfide bonded human insulin was synthesized. The synthesized hormone was identical  
 R:Nathani, V.K.  
 Hoppe-Seyler's Z. Physiol. Chem. 354, 659-672, 1973  
 A:Title: The synthesis of C-peptide of human proinsulin.  
 A:Reference number: A91658; MUID:75040007; PMID:4803404  
 A:Contents: annotation; synthesis of residues 57-87  
 R:Geiger, R.; Jaeger, G.; Koenig, W.  
 Chem. Ber. 106, 2447-2452, 1973  
 A:Title: Synthesis of the complete sequence of human proinsulin C-peptide and its [Glu-9  
 A:Reference number: A90914  
 A:Contents: annotation; synthesis of residue 57-87  
 R:Kaufmann, J.E.; Irminger, J.C.; Halban, P.A.  
 Biochem. J. 110, 869-874, 1995  
 A:Title: Sequence requirements for proinsulin processing at the B-chain/C peptide junction  
 A:Reference number: S58661; MUID:96013185; PMID:7575420  
 A:Contents: annotation; site-directed mutagenesis study of proteolytic processing  
 C:Genetics:  
 A:Gene: GDB:INS  
 A:Cross-references: GDB:119349; OMIM:176730  
 A:Map position: 11p15.3-11p15.5  
 C:Introns: 63/1  
 C:Superfamily: insulin  
 C:Keywords: hormone; pancreas  
 F:1-24/Domain: signal sequence \*status predicted <SIG>  
 F:25-54/Domain: insulin chain B \*status experimental <BCH>  
 F:57-87/Domain: insulin \*status experimental <MAT>  
 F:90-110/Domain: connecting C peptide \*status experimental <CEP>  
 F:31-96, 43-109, 95-100/Disulfide bonds: \*status experimental  
 Query Match 100.0%; Score 463; DB 2; Length 110;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FVNHLGGSHLVEALYLVCGERGFFYTPKTRREAEDLQVQLGGPGAGSLOPLALEG 60  
 DB 25 FVNHLGGSHLVEALYLVCGERGFFYTPKTRREAEDLQVQLGGPGAGSLOPLALEG 84  
 QY 61 SLOKRGIVQCCTSTCSLYOLENYCN 86  
 DB 85 SLOKRGIVQCCTSTCSLYOLENYCN 110  
 RESULT 2  
 A42179  
 Insulin precursor - chimpanzee  
 C:Species: Pan troglodytes (chimpanzee)  
 C:Date: 04-Mar-1993 \*sequence\_revision 18-Nov-1994 \*text\_change 16-Jul-1999  
 C:Accession: A42179; S22058  
 R:Seino, S.; Bell, G.I.; Li, W.H.  
 Mol. Biol. Evol. 9, 193-203, 1992  
 A:Title: Sequences of primate insulin genes support the hypothesis of a slower rate  
 A:Reference number: A42179; MUID:92219453; PMID:1560757  
 A:Accession: A42179  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-110 <SEI>  
 A:Cross-references: EMBL:X61089; NID:q48251; PIDN:CAA43403.1; PID:q48252  
 A:Note: sequence extracted from NCBI backbone (NCBI:95067)  
 C:Genetics:  
 A:Introns: 63/2  
 C:Superfamily: insulin  
 Query Match 100.0%; Score 463; DB 2; Length 110;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FVNHLGGSHLVEALYLVCGERGFFYTPKTRREAEDLQVQLGGPGAGSLOPLALEG 60  
 DB 25 FVNHLGGSHLVEALYLVCGERGFFYTPKTRREAEDLQVQLGGPGAGSLOPLALEG 84  
 QY 61 SLOKRGIVQCCTSTCSLYOLENYCN 86  
 DB 85 SLOKRGIVQCCTSTCSLYOLENYCN 110  
 RESULT 4  
 A42179  
 Insulin precursor - green monkey  
 C:Species: Cercopithecus aethiops (green monkey, aivet)  
 C:Date: 04-Mar-1993 \*sequence\_revision 18-Nov-1994 \*text\_change 16-Jul-1999  
 C:Accession: S42179; A05232; S16494; S22056  
 R:Seino, S.; Bell, G.I.; Li, W.H.  
 Mol. Biol. Evol. 9, 193-203, 1992  
 A:Title: Sequences of primate insulin genes support the hypothesis of a slower rate  
 A:Reference number: A42179; MUID:92219453; PMID:1560757  
 A:Accession: A42179  
 A:Molecule type: DNA  
 A:Residues: 1-110 <SEI>  
 A:Cross-references: EMBL:X61092; NID:q22808; PIDN:CAA43405.1; PID:q22809  
 A:Note: sequence extracted from NCBI backbone (NCBI:95185; NCBI:95194)  
 R:Peterson, J.D.; Nehrllich, S.; Oyer, P.E.; Steiner, D.F.  
 J. Biol. Chem. 247, 4866-4871, 1972  
 A:Title: Determination of the amino acid sequence of the monkey, sheep, and dog proin  
 A:Reference number: A52131; MUID:7258016; PMID:4626469  
 A:Accession: A05232  
 A:Molecule type: protein  
 A:Residues: 57-87 <PEI>  
 C:Genetics:  
 A:Introns: 63/1  
 C:Superfamily: insulin  
 C:Keywords: hormone; pancreas  
 F:1-24/Domain: signal sequence \*status predicted <SIG>  
 F:25-54/Domain: insulin chain B \*status predicted <BCH>  
 F:57-87/Domain: insulin \*status predicted <MAT>  
 F:90-110/Domain: connecting peptide \*status experimental <CEP>  
 F:31-96, 43-109, 95-100/Disulfide bonds: \*status predicted  
 Query Match 98.5%; Score 456; DB 2; Length 110;  
 Best Local Similarity 98.8%; Pred. No. 1.6e-42;  
 Matches 85; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 FVNHLGGSHLVEALYLVCGERGFFYTPKTRREAEDLQVQLGGPGAGSLOPLALEG 60  
 DB 25 FVNHLGGSHLVEALYLVCGERGFFYTPKTRREAEDLQVQLGGPGAGSLOPLALEG 84  
 QY 61 SLOKRGIVQCCTSTCSLYOLENYCN 86  
 DB 85 SLOKRGIVQCCTSTCSLYOLENYCN 110

JQ0178  
 Insulin precursor - crab-eating macaque  
 C:Species: Macaca fascicularis (crab-eating macaque)  
 C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 16-Jul-1999  
 C:Accession: JQ0178  
 R:Wetkam, W.; Gronenberg, J.; Leineweber, M.; Wengenmayer, F.; Winnacker, E.L.  
 Gene 19, 179-183, 1982  
 A:Title: The nucleotide sequence of cDNA coding for preproinsulin from the primate Macaca  
 A:Reference number: JQ0178; MUID:83080474; #text\_change 16-Jul-1999  
 A:Accession: JQ0178  
 A:Molecule type: mRNA  
 A:Residues: 1-110 <MET>  
 A:Cross-references: GB:J00336; NID:9342121; PIDN:AAA36849.1; PID:9342122  
 C:Superfamily: Insulin  
 F:1-24/Domain: signal sequence #status predicted <SIG>  
 F:25-54/Domain: insulin chain B #status predicted <MAT>  
 F:55-84/Domain: insulin chain A #status predicted <ACH>  
 F:85-110/Domain: insulin connecting C-peptide #status predicted <CPEP>  
 F:91-96,43-109,95-100/Disulfide bonds: #status predicted <ACH>  
 F:31-96,43-109,95-100/Disulfide bonds: #status predicted  
 Query Match 98.5%; Score 456; DB 1; Length 110;  
 Best Local Similarity 98.5%; Pred. No. 1; 60-42;  
 Matches 85; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 RESULT 5  
 INRB  
 Insulin precursor - rabbit  
 N:Alternate names: preproinsulin  
 C:Species: Oryctolagus cuniculus (domestic rabbit)  
 C:Date: 24-Apr-1984 #sequence\_revision 23-Aug-1997 #text\_change 18-Jun-1999  
 C:Accession: A53438; A01581  
 R:Devaskar, S.U.; Giddings, S.J.; Rajakumar, P.A.; Carnaghi, L.R.; Menon, R.K.; Zahm, D.  
 J. Biol. Chem. 269, 8445-8454, 1994  
 A:Title: Insulin gene expression and insulin synthesis in mammalian neuronal cells.  
 A:Reference number: A53438; MUID:94179240; PMID:8132571  
 A:Accession: A53438  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-110 <DEV>  
 A:Cross-references: GB:U03610; NID:9467970; PIDN:AAA19033.1; PID:9467971  
 R:Smith, L.F.  
 Am. J. Med. 40, 662-666, 1966  
 A:Title: Species variation in the amino acid sequence of insulin.  
 A:Reference number: A90029; MUID:66160119; PMID:5949593  
 A:Accession: A01581  
 A:Molecule type: protein  
 C:Superfamily: Insulin  
 C:Keywords: hormone; pancreas  
 F:1-24/Domain: signal sequence #status predicted <SIG>  
 F:25-54/Domain: insulin chain B #status experimental <BCH>  
 F:55-84/Domain: insulin chain A #status experimental <MAT>  
 F:85-110/Domain: connecting C-peptide #status predicted <CPEP>  
 F:91-96,43-109,95-100/Disulfide bonds: #status experimental  
 F:31-96,43-109,95-100/Disulfide bonds: #status experimental  
 Query Match 90.1%; Score 417; DB 1; Length 110;  
 Best Local Similarity 89.5%; Pred. No. 2; 9e-38;  
 Matches 77; Conservative 1; Mismatches 8; Indels 0; Gaps 0;  
 RESULT 6  
 IPDG  
 Insulin precursor - dog  
 C:Species: Canis lupus familiaris (dog)  
 C:Date: 24-Apr-1984 #sequence\_revision 15-Nov-1984 #text\_change 16-Jul-1999  
 C:Accession: A92413; A01587; S16493  
 R:Kwok, S.C.M.; Chan, S.J.; Steiner, D.F.  
 J. Biol. Chem. 258, 2357-2363, 1983  
 A:Title: Cloning and nucleotide sequence analysis of the dog insulin gene. Coded am  
 A:Reference number: A92413; MUID:83109071; PMID:6296142  
 A:Accession: A92413  
 A:Molecule type: DNA  
 A:Residues: 1-110 <SMI>  
 A:Cross-references: GB:V30179; GB:100642; NID:94944; PIDN:CAA23475.1; PID:9995  
 R:Smith, L.F.  
 Am. J. Med. 40, 662-666, 1966  
 A:Title: Species variation in the amino acid sequence of insulin.  
 A:Reference number: A90029; MUID:66160119; PMID:5949593  
 A:Accession: A01587  
 A:Molecule type: protein  
 A:Residues: 25-54:90-110 <SMIT>  
 R:Peterson, J.D.; Nehrlich, S.; Oyer, P.E.; Steiner, D.F.  
 J. Biol. Chem. 247, 4866-4871, 1972  
 A:Title: Determination of the amino acid sequence of the monkey, sheep, and dog pro  
 A:Reference number: A92111; MUID:72258016; PMID:4626369  
 A:Accession: S16493  
 A:Molecule type: protein  
 A:Residues: 65-85,17,87 <PET>  
 C:Superfamily: Insulin  
 C:Keywords: hormone; pancreas  
 F:1-24/Domain: signal sequence #status predicted <SIG>  
 F:25-54/Domain: insulin chain B #status experimental <BCH>  
 F:55-84/Domain: insulin chain A #status experimental <MAT>  
 F:85-110/Domain: connecting C-peptide #status predicted <CPEP>  
 F:91-96,43-109,95-100/Disulfide bonds: #status experimental  
 F:31-96,43-109,95-100/Disulfide bonds: #status experimental  
 Query Match 90.1%; Score 417; DB 1; Length 110;  
 Best Local Similarity 89.5%; Pred. No. 2; 9e-38;  
 Matches 77; Conservative 1; Mismatches 8; Indels 0; Gaps 0;  
 RESULT 7  
 IPHO  
 Insulin precursor - horse  
 C:Species: Equus caballus (domestic horse)  
 C:Date: 13-Jul-1981 #sequence\_revision 13-Jul-1981 #text\_change 16-Jul-1999  
 C:Accession: A01580; A92120  
 R:Harris, J.I.; Sanger, F.; Naughton, M.A.  
 Arch. Biochem. Biophys. 65, 427-428, 1956  
 A:Title: Species differences in insulin.  
 A:Reference number: A90082  
 A:Accession: A01580  
 A:Molecule type: protein  
 A:Residues: 1-30:66-86 <HAR>  
 R:Tager, H.S.; Steiner, D.F.  
 J. Biol. Chem. 247, 7936-7940, 1972  
 A:Title: Primary structures of the proinsulin connecting peptides of the rat and ho  
 A:Reference number: A92120; MUID:73061498; PMID:464091

QY 61 SLOKRGIVEQCCTSCSYQLENYCN 86  
 DB 85 ALQKRGIVEQCCTSCSYQLENYCN 110

## RESULT 6

IPDG  
 Insulin precursor - dog  
 C:Species: Canis lupus familiaris (dog)  
 C:Date: 24-Apr-1984 #sequence\_revision 15-Nov-1984 #text\_change 16-Jul-1999  
 C:Accession: A92413; A01587; S16493  
 R:Kwok, S.C.M.; Chan, S.J.; Steiner, D.F.  
 J. Biol. Chem. 258, 2357-2363, 1983  
 A:Title: Cloning and nucleotide sequence analysis of the dog insulin gene. Coded am  
 A:Reference number: A92413; MUID:83109071; PMID:6296142  
 A:Accession: A92413  
 A:Molecule type: DNA  
 A:Residues: 1-110 <SMI>  
 A:Cross-references: GB:V30179; GB:100642; NID:94944; PIDN:CAA23475.1; PID:9995  
 R:Smith, L.F.  
 Am. J. Med. 40, 662-666, 1966  
 A:Title: Species variation in the amino acid sequence of insulin.  
 A:Reference number: A90029; MUID:66160119; PMID:5949593  
 A:Accession: A01587  
 A:Molecule type: protein  
 A:Residues: 25-54:90-110 <SMIT>  
 R:Peterson, J.D.; Nehrlich, S.; Oyer, P.E.; Steiner, D.F.  
 J. Biol. Chem. 247, 4866-4871, 1972  
 A:Title: Determination of the amino acid sequence of the monkey, sheep, and dog pro  
 A:Reference number: A92111; MUID:72258016; PMID:4626369  
 A:Accession: S16493  
 A:Molecule type: protein  
 A:Residues: 65-85,17,87 <PET>  
 C:Superfamily: Insulin  
 C:Keywords: hormone; pancreas  
 F:1-24/Domain: signal sequence #status predicted <SIG>  
 F:25-54/Domain: insulin chain B #status experimental <BCH>  
 F:55-84/Domain: insulin chain A #status experimental <MAT>  
 F:85-110/Domain: connecting C-peptide #status predicted <CPEP>  
 F:91-96,43-109,95-100/Disulfide bonds: #status experimental  
 F:31-96,43-109,95-100/Disulfide bonds: #status experimental  
 Query Match 90.1%; Score 417; DB 1; Length 110;  
 Best Local Similarity 89.5%; Pred. No. 2; 9e-38;  
 Matches 77; Conservative 1; Mismatches 8; Indels 0; Gaps 0;  
 RESULT 7  
 IPHO  
 Insulin precursor - horse  
 C:Species: Equus caballus (domestic horse)  
 C:Date: 13-Jul-1981 #sequence\_revision 13-Jul-1981 #text\_change 16-Jul-1999  
 C:Accession: A01580; A92120  
 R:Harris, J.I.; Sanger, F.; Naughton, M.A.  
 Arch. Biochem. Biophys. 65, 427-428, 1956  
 A:Title: Species differences in insulin.  
 A:Reference number: A90082  
 A:Accession: A01580  
 A:Molecule type: protein  
 A:Residues: 1-30:66-86 <HAR>  
 R:Tager, H.S.; Steiner, D.F.  
 J. Biol. Chem. 247, 7936-7940, 1972  
 A:Title: Primary structures of the proinsulin connecting peptides of the rat and ho  
 A:Reference number: A92120; MUID:73061498; PMID:464091

QY 1 FVNQHLGSHLVEALYLVCGERGFFYTPKTRREAEDLQVQLGGPGAGSLOPLALEG 60  
 DB 25 FVNQHLGSHLVEALYLVCGERGFFYTPKTRREAEDLQVQLGGPGAGSLOPLALEG 84

QY 61 SLOKRGIVEQCCTSCSYQLENYCN 86  
 DB 85 ALQKRGIVEQCCTSCSYQLENYCN 110

## RESULT 7

IPHO  
 Insulin precursor - horse  
 C:Species: Equus caballus (domestic horse)  
 C:Date: 13-Jul-1981 #sequence\_revision 13-Jul-1981 #text\_change 16-Jul-1999  
 C:Accession: A01580; A92120  
 R:Harris, J.I.; Sanger, F.; Naughton, M.A.  
 Arch. Biochem. Biophys. 65, 427-428, 1956  
 A:Title: Species differences in insulin.  
 A:Reference number: A90082  
 A:Accession: A01580  
 A:Molecule type: protein  
 A:Residues: 1-30:66-86 <HAR>  
 R:Tager, H.S.; Steiner, D.F.  
 J. Biol. Chem. 247, 7936-7940, 1972  
 A:Title: Primary structures of the proinsulin connecting peptides of the rat and ho  
 A:Reference number: A92120; MUID:73061498; PMID:464091

QY 1 FVNQHLGSHLVEALYLVCGERGFFYTPKTRREAEDLQVQLGGPGAGSLOPLALEG 60  
 DB 25 FVNQHLGSHLVEALYLVCGERGFFYTPKTRREAEDLQVQLGGPGAGSLOPLALEG 84

A:Accession: A92120  
A:Molecule type: protein  
A:Residues: 33-63 <TAG>  
C:Comment: X's at positions 31-32 and 64-65 represent paired basic residues assumed (by  
C:Superfamily: insulin  
F:1-30/Domain: insulin chain B #status experimental <BCH>  
F:1-30-66-86/Product: insulin #status experimental <MAT>  
F:33-63/Domain: connecting peptide #status experimental <CPP>  
F:66-86/Domain: insulin chain A #status experimental <ACH>  
F:7-72,19-85,71-76/Disulfide bonds: #status predicted

Query Match 85.1%; Score 394; DB 1; Length 86;  
Best Local Similarity 84.9%; Pred. No. 9.2e-36;  
Matches 73; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVFALYLVCGERGFFYTPKTRAEADLVQGVGVGGPGAGSLQPLALEG 60  
DB : FVNHLCGSHLVFALYLVCGERGFFYTPKTRAEADLVQGVGVGGPGAGSLQPLALEG 60  
QY 61 SLQKRGIVFQCCTSCISLYOLENYCN 86  
DB : SLQKRGIVFQCCTSCISLYOLENYCN 86

RESULT 8  
IPRT2  
insulin 2 precursor - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 23-Mar-1981 #sequence,revision: 23-Oct-1981 #text\_change 24-Sep-1999  
C:Accession: B90789; B94231; C92120; I64860; A01590; B92120  
R:Lomedico, P.; Rosenthal, N.; Efstratiadis, A.; Gilbert, W.; Kolodner, R.; Tizard, R.  
Cell 18, 545-558, 1979  
A:Title: The structure and evolution of the two nonallelic rat preproinsulin genes.  
A:Reference number: A90789; MUID:80045035; PMID:458284  
A:Accession: B90789  
A:Molecule type: DNA  
A:Residues: 1-110 <LON>  
A:Cross references: GB:J00748; NID:q204958; PIDN:AAA41443.1; PID:q204959  
R:Steiner, D.F.; Clark, J.L.; Nolan, C.; Rabinstein, A.H.; Margolis, E.; Aten, B.; Oye  
Recent Prot. Form. Res. 25, 207-282, 1969  
A:Title: Proinsulin and the biosynthesis of insulin.  
A:Reference number: A94231; MUID:70067613; PMID:4311938  
A:Accession: B94231  
A:Molecule type: protein  
A:Residues: 25-54;90-110 <STE>  
R:Tager, H.S.; Steiner, D.F.  
J. Biol. Chem. 247, 7936-7940, 1972  
A:Title: Primary structures of the proinsulin connecting peptides of the rat and horse.  
A:Reference number: A92120; MUID:73061498; PMID:4640931  
A:Accession: C92120  
A:Molecule type: protein  
A:Residues: 57-87 <TAG>  
R:Lomedico, P.T.; Rosenthal, N.; Kolodner, R.; Efstratiadis, A.; Gilbert, W.  
Ann. N. Y. Acad. Sci. 343, 425-432, 1980  
A:Title: The structure of rat preproinsulin gene.  
A:Reference number: 151945; MUID:80240379; PMID: 249167  
A:Accession: I64880  
A:Status: preliminary; translated from GB/EMBL/UDBT  
A:Molecule type: DNA  
A:Residues: 1-110 <RES>  
A:Cross references: GB:M25585; NID:q204950; PIDN:AAA41440.1; PID:q204952  
C:Genetics:  
A:Gene: INS2  
A:Introns: 63/1  
C:Superfamily: insulin  
C:Keywords: hormone; pancreas  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:1-24/Domain: insulin chain B #status experimental <BCH>  
F:25-54/Domain: insulin #status experimental <MAT>  
F:57-87/Domain: connecting peptide #status experimental <CPP>  
F:90-110/Domain: insulin chain A #status experimental <ACH>  
F:31-96,43-109,95-100/Disulfide bonds: #status experimental

Query Match 85.1%; Score 394; DB 1; Length 110;  
Best Local Similarity 84.9%; Pred. No. 9.2e-36;  
Matches 73; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVFALYLVCGERGFFYTPKTRAEADLVQGVGVGGPGAGSLQPLALEG 60  
DB : FVNHLCGSHLVFALYLVCGERGFFYTPKTRAEADLVQGVGVGGPGAGSLQPLALEG 60  
QY 61 SLQKRGIVFQCCTSCISLYOLENYCN 86  
DB : SLQKRGIVFQCCTSCISLYOLENYCN 110

RESULT 9  
INS2  
insulin 2 precursor - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 31-Mar-1992 #sequence,revision: 14-Jul-1994 #text\_change 18-Jun-1999  
C:Accession: A26342; B46172; A61012; B01592  
R:Wentworth, B.M.; Schaefer, J.M.; Villa-Komaroff, L.; Chirgwin, J.M.  
J. Mol. Evol. 23, 305-312, 1986  
A:Title: Characterization of the two nonallelic genes encoding mouse preproinsulin.  
A:Reference number: A92565; MUID:87169768; PMID:414503  
A:Accession: A26342  
A:Molecule type: DNA  
A:Residues: 1-110 <MEN>  
A:Cross references: GB:XG4724; NID:q52714; PIDN:CAA28433.1; PID:q52715  
R:Sawa, T.; Ohgaku, S.; Morioka, H.; Yano, S.  
J. Mol. Endocrinol. 5, 61-67, 1990  
A:Title: Molecular cloning and DNA sequence analysis of preproinsulin genes in the N  
A:Reference number: A48172; MUID:90372989; PMID:2397023  
A:Accession: B48172  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-110 <SAW>  
R:Linde, S.; Nielsen, J.H.; Hansen, B.; Wellinder, R.S.  
J. Chromatogr. 462, 243-254, 1989  
A:Title: Reversed-phase high-performance liquid chromatographic analyses of insulin  
A:Reference number: A61012; MUID:89292078; PMID:2661585  
A:Accession: A61012  
A:Molecule type: protein  
A:Residues: 57-87 <LIN>  
R:Buenzli, H.F.; Glatthaar, B.; Kunz, P.; Muelhaupl, E.; Humbel, R.E.  
Hoppe-Sevler's Z. Physiol. Chem. 253, 451-458, 1972  
A:Title: Amino acid sequence of the two insulins from mouse (Mus musculus).  
A:Reference number: A01592; MUID:72189455; PMID:5063718  
A:Accession: B01592  
A:Molecule type: protein  
A:Residues: 25-54;90-110 <BUE>  
C:Genetics:  
A:Introns: 63/1  
C:Superfamily: insulin  
C:Keywords: hormone; pancreas  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-54/Domain: insulin chain B #status experimental <BCH>  
F:25-54,90-110/Product: insulin #status experimental <MAT>  
F:57-87/Domain: connecting peptide #status experimental <CPP>  
F:90-110/Domain: insulin chain A #status experimental <ACH>  
F:31-96,43-109,95-100/Disulfide bonds: #status predicted

Query Match 85.1%; Score 394; DB 1; Length 110;  
Best Local Similarity 84.9%; Pred. No. 9.2e-36;  
Matches 73; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVFALYLVCGERGFFYTPKTRAEADLVQGVGVGGPGAGSLQPLALEG 60  
DB : FVNHLCGSHLVFALYLVCGERGFFYTPKTRAEADLVQGVGVGGPGAGSLQPLALEG 60  
QY 61 SLQKRGIVFQCCTSCISLYOLENYCN 86  
DB : SLQKRGIVFQCCTSCISLYOLENYCN 110

A:Residues: 1-110 <COR>  
A:Cross-references: GB:J00747; NID:q204956; PIDN:AAA41442.1; PID:q204957  
R:Lomedico, P.; Rosenthal, N.; Efstatiadis, A.; Gilbert, W.; Kolodner, R.; Lizard, Cell 18, 545-558, 1979  
A:Title: The structure and evolution of the two nonallelic rat preproinsulin genes.  
A:Reference number: A90789; MUID:80045035; PMID:498284  
A:Accession: A90789  
A:Molecule type: DNA  
A:Residues: 1-110 <LOM>  
A:Cross-references: GB:J00747; NID:q204956; PIDN:AAA41442.1; PID:q204957  
R:Steiner, D.F.; Clark, J.L.; Nolan, C.; Rubenstein, A.H.; Margoliash, E.; Aton, B. Recent Prog. Horm. Res. 25, 207-282, 1969  
A:Title: Proinsulin and the biosynthesis of insulin.  
A:Reference number: A94231; MUID:70067613; PMID:4311938  
A:Accession: A94231  
A:Molecule type: protein  
A:Residues: 25-54; 90-110 <STE>  
R:Taget, H.S.; Steiner, D.F. J. Biol. Chem. 247, 7936-7940, 1972  
A:Title: Primary structures of the proinsulin connecting peptides of the rat and human.  
A:Reference number: A91210; MUID:74561498; PMID:6440951  
A:Accession: A91210  
A:Molecule type: protein  
A:Residues: 57-87 <IAG>  
R:Lomedico, P.T.; Rosenthal, N.; Kolodner, R.; Efstatiadis, A.; Gilbert, W. Ann. N. Y. Acad. Sci. 443, 425-432, 1980  
A:Title: The structure of rat preproinsulin genes.  
A:Reference number: 151945; MUID:80240379; PMID:6249167  
A:Accession: 151945  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-110 <RES>  
A:Cross-references: GB:M25584; NID:q204947; PIDN:AAA41443.1; PID:q204948  
C:Genetics:  
A:Gene: INS1  
C:Superfamily: Insulin  
C:Keywords: hormone; pancreas  
F:1-24/Domain: signal sequence \*status predicted <SIG>  
F:25-54/Domain: insulin chain B \*status experimental <BCB>  
F:54-90-110/Product: insulin \*status experimental <MAT>  
F:57-87/Domain: connecting peptide \*status experimental <CPBP>  
F:90-110/Domain: insulin chain A \*status experimental <ACH>  
F:1-96,43-109,95-100/Disulfide bonds: \*status experimental

Query Match 83.2%; Score 385, DB 1; Length 110;  
Best Local Similarity 83.7%; Pied. No. 8.8e-15;  
Matches 72; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVLEALVLCGGGFFYTKTREAELQVLCVQVLCGGGCAUSLQPLALEG 60  
DB 25 FVQHLCGPHCVLEALVLCGGGFFYTKSRREVDPQVQLELCGGGPGAGLQTLALEV 84  
QY 61 SLQKRGIVQCCSTCSLYOLENYCN 86  
DB 85 AQKRGIVQCCSTCSLYOLENYCN 110

RESULT 13  
IPPC  
insulin precursor - pig  
C:Species: Sus scrofa domestica (domestic pig)  
C:Date: 22-Jun-1981 \*sequence revision 22-Jun-1981 \*text change 16-Jul-1999  
R:Accession: A01583; A94572; S16492; A60835  
R:Chance, R.E.; Ellis, R.M.; Bromer, W.W. Science 161, 165-167, 1968  
A:Title: Porcine proinsulin: characterization and amino acid sequence.  
A:Reference number: A94240; MUID:68286485; PMID:5657063  
A:Accession: A01583  
A:Molecule type: protein  
A:Residues: 1-34, 'Q', 36-84 <CHA>  
R:Chance, R.E.  
submitted to the Atlas, July 1970  
A:Reference number: A94572

A:Accession: A94572  
 A:Molecule type: protein  
 A:Residues: 1-84 <CH2>  
 R:Brown, H.; Sanger, F.; Kitai, R.  
 Biochem. J. 60, 556-565, 1955  
 A:Title: The structure of pig and sheep insulins.  
 A:Reference number: A90344  
 A:Accession: S16492  
 A:Molecule type: protein  
 A:Residues: 1-30, 31-51 <BRO>  
 R:Snell, L.; Damgaard, U.  
 Horm. Metab. Res. 20, 476-480, 1988  
 A:Title: Proinsulin heterogeneity in pigs.  
 A:Reference number: A60835; MUID:89032178; PMID:3181865  
 A:Accession: A60835  
 A:Molecule type: protein  
 A:Residues: 31-38, 40-62 <SNE>  
 A:Note: the authors report the characterization of a connecting peptide variant lacking  
 A:Accession: B60835  
 A:Molecule type: protein  
 A:Residues: 31-42 <SN2>  
 R:Blundell, P.; Dodson, G.; Bolking, P.; Moras, D.  
 Adv. Protein Chem. 26, 279-402, 1972  
 A:Title: Insulin: the structure in the crystal and its reflection in chemistry and biology  
 A:Reference number: A90017  
 A:Accession: A90017  
 A:Contents: annotation: X-ray crystallography, 1.9 angstroms  
 C:Superfamily: insulin  
 C:Keywords: hormone; pancreas  
 F:1-30/Domain: insulin chain B #status experimental <BCH>  
 F:30-64-84/Product: insulin #status experimental <MAT>  
 F:33-63/Domain: connecting peptide #status experimental <CPEP>  
 F:64-84/Domain: insulin chain A #status experimental <ACH>  
 F:7-70, 19-83, 69-74/Disulfide bonds: #status experimental  
 Query Match 82.7% Score 381; DR 1; Length 84;  
 Best Local Similarity 86.0%; Pred. No. 10; 34;  
 Matches 74; Conservative 1; Mismatches 9; Indels 2; Gaps 1;  
 QY 1 FVNHLCGSHLVGALVYVCGERGFYTPK/REAEGLGVGVGGPGAGSLOPLALEG 60  
 Db 1 FVNHLCGSHLVGALVYVCGERGFYTPK/REAEPLVAGAVELG--CLGGLQALALEG 58  
 QY 61 SLKRGIVEQCCTSGCSLYOLENYCN 86  
 Db 59 PPQKRGIVEQCCTSGCSLYOLENYCN 44  
 RESULT 14  
 IPBO  
 Insulin precursor - bovine  
 C:Species: Bos primigenius taurus (cattle)  
 C:Date: 24-Apr-1984 #sequence revision 22-Apr-1995 #text change 16-Jul-1999  
 C:Accession: A40909; A92074; A90342; A90341; S48184; S48185; S46258; A01  
 R:D'Agostino, J.; Yunes, M.A.; White, J.W.; Bosch, P.K.; Field, J.B.; Frazier, H.L.  
 Mol. Endocrinol. 1, 327-331, 1987  
 A:Title: Cloning and nucleotide sequence analysis of complementary deoxyribonucleic acid  
 A:Reference number: A40909; MUID:88268204; PMID:2456452  
 A:Accession: A40909  
 A:Molecule type: mRNA  
 A:Residues: 1-105 <DAA>  
 A:Cross-references: GR-M54979; NID:Q16378; P:DN:AAA30722.1; PID:Q16379  
 A:Experimental source: fetal pancreas  
 R:Nolan, C.; Marcolash, E.; Peterson, J.D.; Steiner, D.F.  
 J. Biol. Chem. 246, 2780-2795, 1971  
 A:Title: The structure of bovine proinsulin.  
 A:Reference number: A92080; MUID:71166442; PMID:4928892  
 A:Accession: A92080  
 A:Molecule type: protein  
 A:Residues: 25-105 <NOI>  
 R:Steiner, D.F.; Cho, S.; Oyer, P.E.; Terris, S.; Peterson, J.D.; Rubenstein, A.H.  
 J. Biol. Chem. 246, 1365-1374, 1971  
 A:Title: Isolation and characterization of proinsulin C-peptide from bovine pancreas.  
 A:Reference number: A92074; MUID:7116409; PMID:5545080

A:Accession: A92074  
 A:Molecule type: protein  
 A:Residues: 57-82 <STE>  
 R:Salokangas, A.; Smyth, D.G.; Markussen, J.; Sundby, F.  
 Eur. J. Biochem. 20, 183-189, 1971  
 A:Title: Bovine proinsulin: amino acid sequence of the C-peptide isolated from pancreas  
 A:Reference number: A91185; MUID:71257721; PMID:5105368  
 A:Accession: A91185  
 A:Molecule type: protein  
 A:Residues: 57-82 <SAL>  
 R:Sanger, F.; Thompson, E.O.P.  
 Biochem. J. 53, 366-374, 1953  
 A:Title: The amino-acid sequence in the diacyl chain of insulin. 2. The investigation  
 A:Reference number: A90342  
 A:Accession: A90342  
 A:Molecule type: protein  
 A:Residues: 85-105 <SAN>  
 R:Sanger, F.; Tuppy, H.  
 Biochem. J. 49, 481-490, 1951  
 A:Title: The amino-acid sequence in the phenylalanyl chain of insulin. 2. The investigation  
 A:Reference number: A90341  
 A:Accession: A90341  
 A:Molecule type: protein  
 A:Residues: 25-54 <SA2>  
 R:Cheng, R.; Kawakishi, S.  
 Eur. J. Biochem. 223, 759-764, 1994  
 A:Title: Site-specific oxidation of histidine residues in glycosylated insulin mediated  
 A:Reference number: S48184; MUID:94333378; PMID:8055951  
 A:Accession: S48184  
 A:Molecule type: protein  
 A:Residues: 85-105 <CHE>  
 A:Accession: S48185  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 25-30, 32-42, 44-54 <CH2>  
 R:Ryle, A.P.; Sanger, F.; Smith, L.F.; Kitai, R.  
 Biochem. J. 60, 541-556, 1955  
 A:Title: The disulphide bonds of insulin.  
 A:Reference number: A90343  
 A:Contents: annotation: amides; disulfides  
 R:Wenzel, T.; Eckerskorn, C.; Lottspeich, F.; Hammerstein, W.  
 FEBS Lett. 349, 205-209, 1994  
 A:Title: Existence of a molecular ruler in proteasomes suggested by analysis of degt  
 A:Reference number: S46258; MUID:94326921; PMID:8050567  
 A:Accession: S46258  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 25-54 <MEN>  
 C:Superfamily: insulin  
 C:Keywords: hormone; pancreas  
 F:1-24/Domain: signal sequence #status predicted <SIG>  
 F:25-54/Domain: insulin chain B #status experimental <BCH>  
 F:25-54, 85-105/Product: insulin #status experimental <MAT>  
 F:57-82/Domain: connecting peptide #status experimental <CPEP>  
 F:85-105/Domain: insulin chain A #status experimental <ACH>  
 F:31-91, 43-104, 90-95/Disulfide bonds: #status experimental  
 Query Match 79.2% Score 366.5; DR 1; Length 105;  
 Best Local Similarity 80.2%; Pred. No. 6; 70-33;  
 Matches 69; Conservative 2; Mismatches 10; Indels 5; Gaps 1;  
 QY 1 FVNHLCGSHLVGALVYVCGERGFYTPK/REAEGLGVGVGGPGAGSLOPLALEG 60  
 Db 25 FVNHLCGSHLVGALVYVCGERGFYTPK/REAEGLGVGVGGPGAGSLOPLALEG 79  
 QY 61 SLKRGIVEQCCTSGCSLYOLENYCN 86  
 Db 80 PPQKRGIVEQCCTSGCSLYOLENYCN 105  
 RESULT 15  
 INMSI  
 Insulin 1 precursor - mouse

Search completed: September 15, 2003, 12:04:19  
Job time : 14.1792 secs

GenCore version 5.1.6  
 Copyright (c) 1993 - 2003 CompuGen Ltd.  
 OM protein - protein search, using sw model  
 Run on: September 15, 2003, 11:54:00 : Search time 8.01434 Seconds  
 (without alignments)  
 504.633 Million cell updates/sec

Title: US-09-423-100-4

Perfect score: 463

Sequence: 1 FVNQHCGLSHLVEALYLVCGL...IVEQCTSIQSLYLENYCN R6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127864

Minimum gap seq length: 0

Maximum DB seq length: 2000000.000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY:ES

Result No.	Score	Query Match	Length	DB ID	Description
1	453	100.0	110	INS_HUMAN	P01308 homo sapien
2	454	100.0	110	INS_PANTR	P30410 pan troglod
3	456	98.5	110	INS_CERAE	P30407 ceropithe
4	456	98.5	110	INS_MACFA	P30406 macaca fasc
5	424	91.6	110	INS_RABIT	P01311 oryctolaqu
6	417	90.1	110	INS_CANFA	P01321 canis fami
7	413	89.2	110	INS_SHEPR	G91X13 spermophilu
8	394	85.1	86	INS_HORSE	P01310 equus cabai
9	194	85.1	110	INS2_MOUSE	P01326 mus musculu
10	194	85.1	110	INS2_RAT	P01323 rattus norv
11	392	84.7	108	INS_AOTTR	P10604 aotus trivi
12	392	84.7	110	INS_CRILA	P01313 cricetus
13	385	83.2	110	INS1_RAT	P01322 rattus norv
14	383	82.7	108	INS_PLIG	P01315 sus scrofa
15	377	81.4	110	INS_PSAOR	G62587 psammomys o
16	366.5	79.2	105	INS_BOVIN	P01317 bos taurus
17	366	79.0	108	INS1_MOUSE	P01325 mus musculu
18	362.5	78.3	105	INS_SHEEP	P01318 ovis aries
19	334.5	72.2	108	INS_RDOSP	P21563 rodentia sp
20	314	67.8	110	INS_CAVPO	P01329 cavia porce
21	277.5	59.9	109	INS_OTDE	P17715 ondon deg
22	276.5	59.7	103	INS_SELRF	P51463 selaphorus
23	265.5	57.3	106	INS2_XENLA	P12707 xenopus lae
24	265.5	57.3	107	INS_CHICK	P01332 gallus gall
25	262.5	56.7	106	INS1_XENLA	P12706 xenopus lae
26	256.5	55.4	51	INS_BALPH	P01312 balaenopter
27	256.5	55.4	51	INS_ELEMA	P01316 elephas max
28	256.5	55.4	81	INS_ANAPL	P01333 anas platyr
29	251.5	54.3	51	INS_ACOCA	P01324 acomys cahi
30	246.5	53.2	51	INS_BALBO	P01314 balaenopter
31	246.5	53.2	51	INS_CAMDR	P01320 camelus dro
32	246.5	53.2	51	INS_CAPHI	P01319 capra hircu
33	245.5	53.0	51	INS_FELCA	P06306 felis silve

34 239.5 51.7 51 1 INS\_DIDMA P18109 didelphis m  
 35 234.5 50.6 51 1 INS\_CHIBR P01327 chinchilla  
 36 231.5 50.0 51 1 INS\_ANGAN P07454 anser anser  
 37 228 49.2 115 1 INS\_VERMO G9W7R2 verasper mo  
 38 227.5 49.1 51 1 INS\_HVSCR P01328 hystrix cri  
 39 227.5 49.1 51 1 INS\_TRASC P13887 trachemys s  
 40 225 48.6 52 1 INS\_ACIGU P81423 acipenser g  
 41 221.5 47.8 116 1 INS\_LOPPZ P01341 lophius pis  
 42 218.5 47.2 51 1 INS\_ORNAN G9QY7 ornithorhyn  
 43 218.5 47.2 105 1 INS\_ONCKE P04467 oncorhynch  
 44 216.5 46.8 89 1 INS\_CALMI P13190 callorhynch  
 45 215 46.4 113 1 INS\_ORENI P81025 oreochromis

# ALIGNMENTS

RESULT 1  
 INS\_HUMAN STANDARD; ERT: 170 AA.  
 AC P01308;  
 DT 21-001-1986 (Ref: 31, created)  
 DT 21-001-1986 (Ref: 31, last sequence update)  
 DT 15 SEP-2003 (Ref: 42, last annotation update)  
 DE Insulin precursor.  
 GN INS.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxID 9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-80120725; PubMed-6243748;  
 RA Bell G.L., Pictet R.L., Rutter W.J., Cordell B., Tischler E.,  
 RA Goodman H.M.;  
 RT "Sequence of the human insulin gene.";  
 RL Nature 284:26-32(1980).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-80216113; PubMed-6248962;  
 RA Ullrich A., Dull T.J., Gray A., Rosius J., Sures I.;  
 RT "Genetic variation in the human insulin gene.";  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-80054779; PubMed-501243;  
 RA Bell G.L., Swain W.F., Pictet R.L., Cordell B., Goodman H.M.,  
 RA Rutter W.J.;  
 RT "Nucleotide sequence of a cDNA clone encoding human preproinsulin.";  
 RL Nature 282:525-527(1979).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-80147417; PubMed-6927840;  
 RA Sures I., Goeddel D.V., Gray A., Ullrich A.;  
 RT "Nucleotide sequence of human preproinsulin complementary DNA.";  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-93364428; PubMed-8358440;  
 RA Lucassen A.M., Bell J.L., Jullier C., Lathrop M.;  
 RT "Susceptibility to insulin dependent diabetes mellitus maps to a 4.1  
 kb segment of DNA spanning the insulin gene and associated VNTR.";  
 RL Nat. Genet. 4:305-310(1993).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-22388257; PubMed-12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Schein C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.L., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

- RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullighy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay E.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren F.J., Lu X., Gibbs R.A.,  
RA Farley J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Blakesley R.W., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Rodriguez A.C., Grimwood J., Green E.D., Dickson M.C.,  
RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smalrus D.E.,  
RA Scherfeld A., Schein J.E., Jones S.J.M., Morra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RL *Proc. Natl. Acad. Sci. U.S.A.* 99:16899-16903(2002).  
RN [7]  
RN SEQUENCE OF 1-59 FROM N.A.  
RN TISSUE-BLOOD:  
RN Fajardy I.L., Weill J.J., Stuckens C.C., Danze P.M.P.;  
RA "Description of a novel RFLP diallelic polymorphism (-127 BspI C/G)  
RT within the 5' region of insulin gene.";  
RL Submitted (JUL-1998) to the EMBL/GenBank/  
RN [8]  
RN SEQUENCE OF 25-54 AND 90-111:  
RA Nicol D.S.B.W., Smith L.F.;  
RT "Amino-acid sequence of human insulin.";  
RL *Nature* 187:483-485(1960).  
RN [9]  
RN SEQUENCE OF 57-87.  
RN MEDLINE-7116410; PubMed-5101771;  
RA Over P.E., Cho S., Peterson J.D., Steh D.F.;  
RT "Studies on human proinsulin. Isolation of amino acid sequence of  
RT the human pancreatic C-peptide.";  
RL *J. Biol. Chem.* 246:1375-1386(1971).  
RN [10]  
RN SEQUENCE OF 57-87.  
RN MEDLINE-7125722; PubMed-5560404;  
RA Ko A., Smyth D.G., Markusen J., Sundby I.;  
RT "The amino acid sequence of the C-peptide of human proinsulin.";  
RL *Eur. J. Biochem.* 20:190-199(1971).  
RN [11]  
RN SYNTHESIS.  
RN MEDLINE-7507277; PubMed-4433293;  
RA Sieber P., Kamber B., Hartmann A., Joehli A., Riniker B., Rittel W.;  
RT "Total synthesis of human insulin under directed formation of the  
RT disulfide bonds.";  
RL *Helv. Chim. Acta* 57:2617-2621(1974).  
RN [12]  
RN SYNTHESIS OF 57-87.  
RN MEDLINE-75040007; PubMed-4803504;  
RA Naithani V.K.;  
RT "Studies on polypeptides. IV. The synthesis of C-peptide of human  
RT proinsulin.";  
RL Hoppe-Seyler's Z. Physiol. Chem. 354:659-672(1973).  
RN [13]  
RN SYNTHESIS OF 65-69 AND 70-73  
RN MEDLINE-73161263; PubMed-4698555;  
RA Geiger P., Volk A.;  
RT "Synthesis of peptides with the properties of human proinsulin C  
RT peptides (hC peptide). 3. Synthesis of the sequences 14-17 and 9-13  
RT of human proinsulin C-peptides.";  
RL *Chem. Ber.* 106:199-205(1973).  
RN [14]  
RN SYNTHESIS OF 84-87.  
RN MEDLINE-73161261; PubMed-4698551;  
RA Geiger P., Jaeger G., Keonig W., Treutlein G.;  
RT "Synthesis of peptides with the properties of human proinsulin C  
RT peptides (hC peptide). 1. Scheme for the synthesis and preparation of  
RT the sequence 28-31 of human proinsulin C peptide.";  
RL *Chem. Ber.* 106:188-192(1973).  
RN [15]  
RN VARIANT LOS ANGELES SER-48.  
RN MEDLINE-84016053; PubMed-6312455;  
RA Haneda M., Chan S.-J., Kwok S.C.M., Rubenstein A.H., Steiner D.F.;  
RT "Studies on mutant human insulin genes: identification and sequence  
RT analysis of a gene encoding [SerB24]insulin.";  
RL *Proc. Natl. Acad. Sci. U.S.A.* 80:6366-6370(1983).  
RN [16]  
RN VARIANTS LOS ANGELES SER-48 AND CHICAGO LEU-49.  
RN MEDLINE-84170233; PubMed-6424111;  
RA Shoelson S., Fickova M., Haneda M., Nahum G., Kaiser E.T.,  
RA Rubenstein A.H., Tager H.;  
RT "Identification of a mutant human insulin predicted to contain a  
RT serine-for-phenylalanine substitution.";  
RL *Proc. Natl. Acad. Sci. U.S.A.* 80:7390-7394(1983).  
RN [17]  
RN VARIANT PROVIDENCE ASP 34.  
RN MEDLINE-87175640; PubMed-3470784;  
RA Chan S.-J., Seino S., Gruppuso P.A., Schwartz R., Steiner D.F.;  
RT "A mutation in the B chain coding region is associated with impaired  
RT proinsulin conversion in a family with hyperproinsulinemia.";  
RL *Proc. Natl. Acad. Sci. U.S.A.* 84:2194-2197(1987).  
RN [18]  
RN VARIANT WAKAYAMA LEU 92.  
RN MEDLINE-87058122; PubMed-3537011;  
RA Sakura H., Iwamoto Y., Sakamoto Y., Kuzuya T., Hirata H.;  
RT "Structurally abnormal insulin in a diabetic patient. Characterization  
RT of the mutant insulin A3 (Val-->Leu) isolated from the pancreas.";  
RL *J. Clin. Invest.* 78:1466-1472(1986).  
RN [19]  
RN VARIANT HIS-89.  
RN MEDLINE-90317021; PubMed-2196279;  
RA Barbetti F., Raben N., Kadowaki T., Cama A., Accili D., Gabbay K.H.,  
RA Merenich J.A., Taylor S.L., Roth J.;  
RT "Two unrelated patients with familial hyperproinsulinemia due to a  
RT mutation substituting histidine for arginine at position 65 in the  
RT proinsulin molecule: identification of the mutation by direct  
RT sequencing of genomic deoxyribonucleic acid amplified by polymerase  
RT chain reaction.";  
RL *J. Clin. Endocrinol. Metab.* 71:164-169(1990).  
RN [20]  
RN VARIANT HIS-89.  
RN MEDLINE-85261996; PubMed-4019786;  
RA Shibasaki Y., Kawakami T., Kanazawa Y., Akanuma Y., Takaku F.;  
RT "Posttranslational cleavage of proinsulin is blocked by a point  
RT mutation in familial hyperproinsulinemia.";  
RL *J. Clin. Invest.* 76:378-380(1985).  
RN [21]  
RN VARIANT KYOTO LEU-89.  
RN MEDLINE-92291307; PubMed-1601997;  
RA Yano H., Kitano N., Morimoto M., Polonsky K.S., Imura H., Seino Y.;  
RT "A novel point mutation in the human insulin gene giving rise to  
RT hyperproinsulinemia (proinsulin Kyoto).";  
RL *J. Clin. Invest.* 89:1902-1907(1992).  
RN [22]  
RN STRUCTURE BY NMR.  
RN MEDLINE-91104966; PubMed-2271664;  
RA Hua Q.-X., Weiss M.A.;  
RT "Toward the solution structure of human insulin: sequential 2D NMR  
RT assignment of a des-pentapeptide analogue and comparison with crystal  
RT structure.";  
RL *Biochemistry* 29:10545-10555(1990).  
RN [23]  
RN STRUCTURE BY NMR.  
RN MEDLINE-91242467; PubMed-2036420;  
RA Hua Q.-X., Weiss M.A.;  
RT "Comparative 2D NMR studies of human insulin and des-pentapeptide  
RT insulin: sequential resonance assignment and implications for protein  
RT dynamics and receptor recognition.";  
RL *Biochemistry* 30:5505-5515(1991).  
RN [24]  
RN STRUCTURE BY NMR.  
RN MEDLINE-91265527; PubMed-1646635;  
RA Hua Q.-X., Weiss M.A.;  
RT "Two-dimensional NMR studies of des-(B26-B30)-insulin: sequence-  
RT specific resonance assignments and effects of solvent composition.";  
RL *Biochim. Biophys. Acta* 1078:101-110(1991).

1 FYNQHLQASHLVLEALYLCVCGEGFFYTPKTRAEADIQVGVGLGGPGAGSLQPLALEG 60  
 25 FYNQHLGGSHLVLEALYLCVCGEGFFYTPKTRAEADIQVGVGLGGPGAGSLQPLALEG 84  
 61 SLOKRGIVQEQCTCSICSLYQLENYCN 86  
 85 SLOKRGIVQEQCTCSICSLYQLENYCN 110

RESULT 3  
 ID INS\_CERAF STANDARD: PRT: 110 AA.  
 AC P30407; P01309;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE Insulin precursor.  
 GN INS.

Caracal: Felis t. aethiops (Green monkey) (Gibbel).  
 Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:  
 Mammalia: Eutheria: Primates: Catarrhini: Ferocepithecidae:  
 Ceropithecinae: Ferocepithecus.  
 NCBI TaxID 9534;  
 111  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE: 92219953; PubMed: 1560757;  
 RX Seino S., Bell G.I., Li W.;  
 RA "Sequences of primate insulin genes support the hypothesis of a  
 RT slower rate of molecular evolution in humans and apes than in  
 RI monkeys.";  
 RI Mol. Biol. Evol. 9:193-203(1992).  
 RN [2]  
 RP SEQUENCE OF 57-87.  
 RX MEDLINE: 72258016; PubMed: 4626369;  
 RA Peterson J.D., Nehrlich S., Uyer P.E., Steiner D.F.;  
 RT "Determination of the amino acid sequence of the monkey, sheep, and  
 RI dog proinsulin C-peptides by a semi-micro Edman degradation  
 RI procedure.";  
 RI J. Biol. Chem. 247:4866-4871(1972).  
 RL  
 CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
 CC  
 CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
 CC DISULFIDE BONDS.  
 CC  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announcement)  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

-----  
 EMBL: X61092; CAA43405.1;  
 PIR: F42179; F42179.  
 HSSP: P01308; A1U.  
 InterPro: IPR004825; Ins/IGF/relax.  
 Pfam: PF00049; Insulin; 1.  
 SMART: SM00078; IIGF; 1.  
 PROSITE: PS00262; INSULIN; 1.  
 Insulin family; Hormone; Glucose metabolism; Signal.  
 SIGNAL: 1 24  
 FT CHAIN 25 54 INSULIN B CHAIN.  
 FT PROPEP 57 87 C PEPTIDE.  
 FT CHAIN 90 110 INSULIN A CHAIN.  
 FT DISULFID 31 95 INTERCHAIN.  
 FT DISULFID 43 109 INTERCHAIN.  
 FT DISULFID 95 100  
 FT SEQUENCE 110 AA: 12019 MW: 95A1F54BE7B247F9 CRC64:

Query Match 98.5%; Score 456; DB 1; Length 110;  
 Best Local Similarity 98.8%; Pred. No. 5.2e-42;  
 Matches 85; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVEALYLVCGRGFFYTPKTRAEADLVQGVELGGPGAGSLCPALLEG 60  
 DB 25 FVNHLCGSHLVEALYLVCGRGFFYTPKTRAEADLVQGVELGGPGAGSLCPALLEG 84  
 OY 61 SLQKRGIVQCCTSCSLYLENYCN 86  
 DB 85 SLQKRGIVQCCTSCSLYLENYCN 110

RESULT 4

INS\_MACFA STANDARD: PRT: 110 AA.

ID INS\_MACFA  
 AC P01309;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE Insulin precursor.  
 GN INS.  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Insuperioria; Leporidae; Oryzologus.  
 OX NCBI\_TaxID:9543;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE 81080474; PubMed:6184262;  
 RA Wetekamp W., Gronenberg J., Leineweber M., Wenckenmayer F.,  
 RA Wisnack E.L.;  
 RT \*The nucleotide sequence of cDNA coding for preproinsulin from the  
 RT primate Macaca fascicularis.\*  
 RL Gene 19:179-183(1982).  
 CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
 CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
 CC DISULFIDE BONDS.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: J00336; AAA56849.1;  
 DR PIR: J00378; J00178.  
 DR HSSP: P01308; 1A10.  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1.  
 DR PRINTS: PR00277; INSULINR.  
 DR SMART: SM00076; IIGF; 1.  
 DR PROSITE: PS00262; INSULIN; 1.  
 KW Insulin family; Hormone; Glucose metabolism; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 54  
 FT PROPEP 57 87  
 FT CHAIN 90 110  
 FT CHAIN 90 110

INSULIN B CHAIN  
 C PEPTIDE  
 INSULIN A CHAIN

Matches 85; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVEALYLVCGRGFFYTPKTRAEADLVQGVELGGPGAGSLCPALLEG 60  
 DB 25 FVNHLCGSHLVEALYLVCGRGFFYTPKTRAEADLVQGVELGGPGAGSLCPALLEG 84  
 OY 61 SLQKRGIVQCCTSCSLYLENYCN 86  
 DB 85 SLQKRGIVQCCTSCSLYLENYCN 110

RESULT 5

INS\_RAHIT STANDARD: PRT: 110 AA.

ID INS\_RAHIT  
 AC P01311;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Insulin precursor.  
 GN INS.  
 OS Oryzologus curvicaulis (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Insuperioria; Leporidae; Oryzologus.  
 OX NCBI\_TaxID:9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-New Zealand white; TISSUE: Pancreas;  
 RX MEDLINE 94179230; PubMed:8142571;  
 RA Devaskar S.U., Giddings S.J., Rajakumar P.A., Carnaghi L.R.,  
 RA Mezon R.K., Zahm D.S.;  
 RT \*Insulin gene expression and insulin synthesis in mammalian neuronal  
 RT cells.\*  
 RL J. Biol. Chem. 259:8445-8454(1994).  
 RN [2]  
 RP SEQUENCE OF 25-54 AND 90-110.  
 RX MEDLINE 66160119; PubMed:5949593;  
 RA Smith L.F.;  
 RT \*Species variation in the amino acid sequence of insulin.\*  
 RL Am. J. Med. 40:662-666(1966).  
 RN [3]  
 RP SEQUENCE OF 56-110 FROM N.A.  
 RA Giddings S.J., Carnaghi L.R., Devaskar S.U.;  
 RL Submitted (Apr-1991) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
 CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
 CC DISULFIDE BONDS.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: U03610; AAA19044.1;  
 DR EMBL: M61154; AAA17540.1;  
 DR PIR: A53438; INRB.  
 DR HSSP: P01308; 1TYM.  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1.  
 DR SMART: SM00076; IIGF; 1.  
 DR PROSITE: PS00262; INSULIN; 1.  
 KW Insulin family; Hormone; Glucose metabolism; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 54  
 FT PROPEP 57 87  
 FT CHAIN 90 110  
 FT CHAIN 90 110

INSULIN B CHAIN  
 C PEPTIDE  
 INSULIN A CHAIN

```

FT PROPEP      57      87      C PEPTIDE.
FT CHAIN       90     110    INSULIN A CHAIN.
FT DISULFID   31      96    INTERCHAIN.
FT FT          43      109    INTERCHAIN.
FT DISULFID   95     100
SQ SEQUENCE   110 AA; 12190 MW;  A574791864A4FB98 CRC64;

Query Match      90.1%; Score 417; DB 1; Length 110;
Best Local Similarity 89.5%; Pred. No. 7.7e-38;
Matches 77; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

OY 1 FVQHLCGSHLVATLYLVCGERGFYTPKTRRAEDUQGVQLGGCPGAGSLQPLALEG 60
Db 25 FVQHLCGSHLVATLYLVCGERGFYTPKTRRAEDUQGVQLGGCPGAGSLQPLALEG 60
OY 61 SLCKRGIVEOCCTSIQSILYQENYN 86
Db 85 ALCKRGIVEOCCTSIQSILYQENYN 110

RESULT 7
INS_SPEAR 1D INS_SPEEK SEAN-LAND; ERT: 110 AA.
AC QY:X4;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE DE INSulin precursor
GN INS.
OS Spermophilus tridecemlineatus (Thirteen-lined ground squirrel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;
OC Spermophilus.
OC NCBI_TaxId=43179;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Tredrea M.M., Buck M.J., Guhaniyoyi J., Squire T.L., Andrews M.T.;
RI "Regulation of Puck M4 expression in a hibernating mammal.";
RI submitted (JUN-2001) to the FMBI/GenBank/DBJ databases.
CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: AY038604; AAK72558.1; ..
HSSP: P01308; 1LNP.
InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; insulin; 1.
DR PRINTS: PR00277; INSULIN.
DR SMART: SM00078; ILGF_1.
DR PROSITE: PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism; Signal.
FT SIGNAL : 24 BY SIMILARITY.
FT CHAIN : 25 54 INSULIN B CHAIN.
FT PROPEP : 57 87 C PEPTIDE.
FT CHAIN : 90 110 INSULIN A CHAIN.
FT DISULFID : 31 96 INTERCHAIN (BY SIMILARITY).
FT DISULFID : 43 109 INTERCHAIN (BY SIMILARITY).
FT DISULFID : 95 100 BY SIMILARITY.
SQ SEQUENCE 110 AA; 12004 MW; 45117680662BEE5 CRC64;

```

Query Match 89.2% Score 413; DB 1; Length 110;  
 Best Local Similarity 89.5%; Pred. No. 20-37;  
 Matches 77; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLYEALVLCGERGFFYTPKTRREAEALQVGVGGPGAGSLQPLALEG 60  
 DB 25 FVNHLCGSHLYEALVLCGERGFFYTPKSRREVEEQGGVGGPGAGLQPLALEM 84  
 QY 61 SLOKRGIVEOCCTSCISLYOLENYCN 86  
 DB 85 ALOKRGIVEOCCTSCISLYOLENYCN 110

RESULT 8  
 INS\_HORSE STANDARD; PKT; 86 AA.  
 AC P01310;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DE 15-JUL-1989 (Rel. 38, Last annotation update)  
 DE Insulin precursor  
 GN INS  
 OS Equus caballus (horse)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Perissodactyla; Equi; Equus  
 OX NCBI\_TaxID=9796;  
 RN [1]  
 RP SEQUENCE OF 1-30 AND 56-86.  
 RA Harris J.I., Sanger F., Naughton M.A.;  
 RT "Species differences in insulin."  
 RL Arch. Biochem. Biophys. 65:427-438(1966)  
 RN [2]  
 RP SEQUENCE OF 38-63.  
 RX MEDLINE 73061498; PubMed=4640931;  
 RA Tager H.S., Steiner D.F.;  
 RT "Primary structures of the proinsulin connecting peptides of the rat  
 and the horse."

CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
 INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
 FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
 CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
 CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
 DISULFIDE BONDS.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 CC -1- CAUTION: X'S AT POSITIONS 31-32 AND 64-65 REPRESENT PAIRED BASIC  
 RESIDUES ASSUMED BY HOMOLOGY TO BE PRESENT IN THE PRECURSOR  
 MOLECULE.

DR PIR: A01580; IPIQ.  
 DR HSSP: P01317; IAPH.  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1.  
 DR SMART: SM00078; IIGF; 1.  
 DR PROSITE: PS00262; INSULIN; 1.  
 KW Insulin family; Hormone; Glucose metabolism.  
 FT CHAIN 1 70 INSULIN B CHAIN.  
 FT PROPEP 33 63 C PEPTIDE.  
 FT CHAIN 66 86 INSULIN A CHAIN.  
 FT DISULFID 7 72 INTERCHAIN.  
 FT DISULFID 19 85 INTERCHAIN.  
 FT DISULFID 71 76 INTERCHAIN.

Query Match 85.1% Score 394; DB 1; Length 86;  
 Best Local Similarity 84.9%; Pred. No. 17e-35;  
 Matches 73; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLYEALVLCGERGFFYTPKTRREAEALQVGVGGPGAGSLQPLALEG 60  
 DB 1 FVNHLCGSHLYEALVLCGERGFFYTPKAXEAEADQVGVGGPGAGLQPLALAG 60

QY 61 SLOKRGIVEOCCTSCISLYOLENYCN 86  
 DB 61 P00XGIVEOCCTSCISLYOLENYCN 86

RESULT 9  
 INS2\_MOUSE STANDARD; PKT; 110 AA.  
 AC P01326;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Insulin 2 precursor.  
 GN INS2 OR INS-2.  
 OS Mus musculus (mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA Knebelworth B.M., Schaffer J.M., Villa-Komaroff L., Chirwin J.M.;  
 RT "Characterization of the two nonallelic genes encoding mouse  
 preproinsulin."

RL J. Mol. Evol. 23:405-412(1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RA Sava T., Ohaku S., Morioka H., Yano S.;  
 RT "Molecular cloning and DNA sequence analysis of preproinsulin genes  
 in the NON mouse, an animal model of human non-obese, non-insulin-

dependent diabetes mellitus."

RL J. Mol. Endocrinol. 5:61-67(1990).  
 RN [3]  
 RP SEQUENCE OF 25-54 AND 90-110.

RA Buendia H.F., Glatthar B., Kunz P., Muelhaupt E., Humbel R.E.;  
 RT "Amino acid sequence of the two insulins from mouse (Maus musculus)."

RL Hoppe-Seyler's Z. Physiol. Chem. 353:451-458(1972).  
 CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
 INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
 FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
 CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.

CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
 DISULFIDE BONDS.

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X04724; CAA28433.1;  
 DR PIR: A26342; INMS2

DR HSSP: P01317; IAPH.

DR MG: MG196573; Ins2.

DR GO: GO:0000187; P:activation of MAPK; IDA.

DR GO: GO:0042325; P:regulation of phosphorylation; IDA.

DR InterPro: IPR004825; Ins/IGF/relax.

DR Pfam: PF00049; Insulin; 1.

DR SMART: SM00078; IIGF; 1.

DR PROSITE: PS00262; INSULIN; 1.

DR Insulin family; Hormone; Glucose metabolism; Signal; Multigene family.

FT SIGNAL 1 24

FT CHAIN 25 54 INSULIN 2 B CHAIN.

FT PROPEP 57 87 INSULIN 2 C PEPTIDE.

FT CHAIN 90 110 INSULIN 2 A CHAIN.

FT DISULFID 31 96 INTERCHAIN.

```

FT DISULFID 43 109 INTERCHAIN.
FT DISULFID 95 100
SQ SEQUENCE 110 AA: 12364 MW: 3554C8803D24FDAD CRC64:

Query Match 85.1% Score 394; DB 1; Length 110;
Best Local Similarity 84.9%; Pred. No. 2.2e-35;
Matches 73; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLYEALYLVCGERGFFYTKTREAEDELQVQVLEGGPCAGSLQPIALEG 60
DB 25 FVKHLCGSHLYEALYLVCGERGFFYTKTREAEDELQVQVLEGGPCAGSLQPIALEG 84
QY 61 SLOKRGIVQCCCTSCSLYQLENYCN 86
DB 85 AOKRGIVDQCCCTSCSLYQLENYCN 110

RESULT 10
INS2_RAT STANDARD; PRT: 119 AA.
AC P01424.
DT 21 JUL 1986 (Rel. 01, Created)
DT 21 JUL 1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Insulin 2 precursor.
GN INS2 OR INS-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurora; Muridae; Murinae; Rattus.
OX NCBI_TaxID:10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:Sprague-Dawley; Tissue:Liver;
RX MEDLINE:86045035; PubMed:498284;
RA Lomedico P., Rosenthal N., Efratiadis A., Gilbert W., Kolodner R.,
RA Tizard R.;
RT "The structure and evolution of the two nonallelic rat preproinsulin
RT genes.";
RL Cell 18:545-558(1979).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE:86310882; PubMed:2427970;
RA Seares M.B., Schin E., Henderson A., Karathanasis S.K., Cate R.,
RA Zeitlin S., Chirwin J., Efratiadis A.;
RT "RNA-mediated gene duplication: the rat preproinsulin I gene is a
RT functional retroposon.";
RL Mol. Cell. Biol. 5:2090-2103(1985).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE:80240379; PubMed:6249167;
RA Lomedico P.T., Rosenthal N., Kolodner R., Efratiadis A.,
RA Gilbert W.;
RT "The structure of rat preproinsulin genes.";
RL Ann. N.Y. Acad. Sci. 343:425-432(1980).
RN [4]
RP SEQUENCE OF 25-54 AND 90-110.
RX MEDLINE:70067613; PubMed:4111918;
RA Steiner D.F., Clark J.L., Nolan C., Kubanstein A.H., Margoliash E.,
RA Aten B., Oyer P.E.;
RT "Proinsulin and the biosynthesis of insulin.";
RL Recent Prog. Horm. Res. 25:207-242(1969).
RN [5]
RP SEQUENCE OF 57-87.
RX MEDLINE:73061498; PubMed:4640931;
RA Tager H.S., Steiner D.F.;
RT "Primary structures of the proinsulin connecting peptides of the rat
RT and the horse.";
RL J. Biol. Chem. 247:7936-7940(1972).
RN [6]
RP SEQUENCE OF 57-87, AND REVISIONS.
RX MEDLINE:72177385; PubMed:4554104;
RA Markussen J., Sundby F.;
RT "Rat-proinsulin C-peptides. Amino-acid sequences.";
```

```

RL Bur. J. Biochem. 25:153-162(1972).
CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DB EMBL: V01243; CAA24560.1;
DB EMBL: J00748; AAA1443.1;
DB EMBL: M25585; AAA1443.1;
DB EMBL: M25583; AAA1443.1; JOINED
DB F18; H95789; 18572
DB HSSP: P01317; IAPH
DR InterPro: IPR014825; Ins/IGF/relex.
DR Pfam: PF00049; Insulin; 1.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism; Signal; Multiqtene family.
FT SIGNAL 1 24
FT CHAIN 25 54 INSULIN 2 B CHAIN
FT PROPEP 57 87 INSULIN 2 C PEPTIDE.
FT CHAIN 96 110 INSULIN 2 A CHAIN.
FT DISULFID 31 96 INTERCHAIN.
FT DISULFID 43 109 INTERCHAIN.
FT DISULFID 95 100
SQ SEQUENCE 110 AA: 12339 MW: 34626DA98C86F3CA CRC64:

Query Match 85.1% Score 394; DB 1; Length 110;
Best Local Similarity 84.9%; Pred. No. 2.2e-35;
Matches 73; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLYEALYLVCGERGFFYTKTREAEDELQVQVLEGGPCAGSLQPIALEG 60
DB 25 FVKHLCGSHLYEALYLVCGERGFFYTKTREAEDELQVQVLEGGPCAGSLQPIALEG 84
QY 61 SLOKRGIVQCCCTSCSLYQLENYCN 86
DB 85 AOKRGIVDQCCCTSCSLYQLENYCN 110

RESULT 11
INS_AOTIR STANDARD; PRT: 108 AA.
AC P10604.
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Insulin precursor.
GN INS.
OS Aotus trivirgatus (Night monkey) (Douroucoulis), and
OS Saimiri sciureus (Common squirrel monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Aotinae; Aotus.
OX NCBI_TaxID:9505; 9521;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES-A. trivirgatus;
RX MEDLINE:88041119; PubMed:3118367;
RA Seino S., Steiner D.F., Bell G.I.;
RT "Sequence of a New World primate insulin having low biological
RT potency and immunoreactivity.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:7423-7427(1987).
```

RN RP  
 RC SPECIES: S. sciureus;  
 RX MEDLINE: 91088593; PubMed: 2263627;  
 RA Yu J.-H., Eng J., Yalow R.S.;  
 RT "Isolation and amino acid sequences of a purified monkey (Saimiri  
 RL sciurea) insulin and glucagon."  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9766-9768(1990).  
 CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
 CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
 CC DISULFIDE BONDS.  
 CC -1- SURCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL: M26328; AAA37089.1;  
 DR HSSP: P01308; IITYM;  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1;  
 DR SMART: SM00078; IIGF; 1;  
 DR PROSITE: PS00262; INSULIN; 1;  
 KW Insulin family; Hormone; Glucose metabolism; Signal.  
 FT SIGNAL: 1 24  
 FT CHAIN: 25 54 INSULIN B CHAIN.  
 FT PROPEP: 57 85 C PEPTIDE.  
 FT CHAIN: 98 108 INSULIN A CHAIN.  
 FT DISULFID: 31 94 INTERCHAIN.  
 FT DISULFID: 43 107 INTERCHAIN.  
 FT DISULFID: 93 98  
 SQ SEQUENCE 108 AA: 1869RRZ50G9973IF CRC64:  
 Query Match 84.7%; Score 492; DR 1; Length 108;  
 Best Local Similarity 84.9%; Pred. No. 3, 5, 35;  
 Matches 73; Conservative 4; Mismatches 7; Indels 2; Gaps 1;  
 QY 1 FVNHGCGSHLVEALYLVCGERGFFYTPKTRPAEDILQGVGVLEGGSPAGSLQALEG 60  
 DB 25 FVNHGCGSHLVEALYLVCGERGFFYTPKTRPAEDILQGVGVLEGGSPAGSLQALEG 84  
 QY 61 SLQKRGIVRQCCTISIGSLYOLENYCN 86  
 DB 85 AOKKRGIVDQCCTISIGSLYOLENYCN 110  
 RESULT 12  
 ID INS\_CR100 STANDARD; PRT: 1; AA  
 AC P01308;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE Insulin precursor.  
 GN INS.  
 OS Cricetus longicaudatus (Long-tailed hamster) (Chinese hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 CC Cricetulus.  
 OX NCBI\_TaxID=10030;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 84133036; PubMed: 6365663;

RA Bell G.I., Sanchez-Pescador R.;  
 RT "Sequence of a cDNA encoding Syrian hamster preproinsulin."  
 RL Diabetes 33:297-300(1984).  
 RN [2]  
 RP SEQUENCE OF 25-54 AND 90-110.  
 RA Neelon F.A., Delcher H.K., Steinman H., Lebovitz H.E.;  
 RT "Structure of hamster insulin: comparison with a tumor insulin."  
 RL Fed. Proc. 32:300-300(1973).  
 CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
 CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
 CC DISULFIDE BONDS.  
 CC -1- SURCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL: M26328; AAA37089.1;  
 DR HSSP: P01308; IITYM;  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1;  
 DR SMART: SM00078; IIGF; 1;  
 DR PROSITE: PS00262; INSULIN; 1;  
 KW Insulin family; Hormone; Glucose metabolism; Signal.  
 FT SIGNAL: 1 24  
 FT CHAIN: 25 54 INSULIN B CHAIN.  
 FT PROPEP: 57 85 C PEPTIDE.  
 FT CHAIN: 98 110 INSULIN A CHAIN.  
 FT DISULFID: 31 96 INTERCHAIN.  
 FT DISULFID: 43 109 INTERCHAIN.  
 FT DISULFID: 95 103  
 SQ SEQUENCE 110 AA: 219E52B85A535GEC CRC64:  
 Query Match 84.7%; Score 492; DR 1; Length 110;  
 Best Local Similarity 84.9%; Pred. No. 3, 5, 35;  
 Matches 73; Conservative 4; Mismatches 9; Indels 0; Gaps 0;  
 QY 1 FVNHGCGSHLVEALYLVCGERGFFYTPKTRPAEDILQGVGVLEGGSPAGSLQALEG 60  
 DB 25 FVNHGCGSHLVEALYLVCGERGFFYTPKTRPAEDILQGVGVLEGGSPAGSLQALEG 84  
 QY 61 SLQKRGIVRQCCTISIGSLYOLENYCN 86  
 DB 85 AOKKRGIVDQCCTISIGSLYOLENYCN 110  
 RESULT 13  
 ID INS1\_RAT STANDARD; PRT: 110 AA.  
 AC P01322;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Insulin 1 precursor.  
 GN INS1 OR INS-1  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 80045034; PubMed: 498283;  
 RA Cordell B., Bell G.I., Tischer E., Denoto F.M., Ullrich A.,  
 RA Picot R.L., Rutter W.J., Goodman H.M.;  
 RT "Isolation and characterization of a cloned rat insulin gene."

Cell 18:533-543(1979).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-Sprague-Dawley; TISSUE=Liver;  
 RX MEDLINE=80045035; PubMed=498284;  
 RA Lomedico P., Rosenthal N., Efratiadis A., Gilbert W., Kolodner R.,  
 RT Tizard R.;  
 RT "The structure and evolution of the two nonallelic rat preproinsulin  
 genes.";  
 RL Cell 18:545-558(1979).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=80240179; PubMed=6249167;  
 RA Lomedico P.T., Rosenthal N., Kolodner R., Efratiadis A.,  
 RT Gilbert W.;  
 RT "The structure of rat preproinsulin genes.";  
 RL Ann. N.Y. Acad. Sci. 343:425-432(1980).  
 RN [4]  
 RP SEQUENCE OF 25-54 AND 90-110.  
 RX MEDLINE=70067613; PubMed=411948;  
 RA Steiner D.F., Clark J.L., Nollan G., Eidenstein A.H., Mironiash E.,  
 RT Allen B., Byers P.E.;  
 RT "Proinsulin and the biosynthesis of insulin.";  
 RL Revet. Crit. Rev. Biochem. Mol. Biol. 25:207-282(1990).  
 RN [5]  
 RP SEQUENCE OF 67-87  
 RX MEDLINE=70061498; PubMed=4040931;  
 RA Tager H.S., Steiner D.F.;  
 RT "Primary structures of the proinsulin connecting peptides of the rat  
 and the horse.";  
 RL J. Biol. Chem. 247:7936-7940(1972).  
 RN [6]  
 RP SEQUENCE OF 57-87, AND REVISIONS.  
 RX MEDLINE=72177385; PubMed=4554104;  
 RA Markussen J., Sundby F.;  
 RT "Rat proinsulin C-peptides. Amino-acid sequences.";  
 RL Eur. J. Biochem. 25:153-162(1972).  
 CC [7] FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
 CC [8] SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
 CC DISULFIDE BONDS.  
 CC [9] SUBCELLULAR LOCATION: Secreted.  
 CC [10] SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: V01242; CAA24559.1; .  
 DR EMBL: J00747; AAA1442.1; .  
 DR EMBL: M25584; AAA1439.1; .  
 DR PIR: A90788; IPRT1.  
 DR HSP: P01308; IATF.  
 DR InterPro: IPR004825; Ins/IGF/relax  
 DR Pfam: PF00349; Insulin; 1.  
 DR SMART: SM00078; IIGF; 1.  
 DR PROSITE: PS00262; INSULIN; 1.  
 KW Insulin family; Hormone; Glucose metabo ism; Signal; Multigene family.  
 FT SIGNAL 1 24  
 FT CHAIN 25 54 INSULIN 1 B CHAIN.  
 FT PROPEP 57 87 INSULIN 1 C PEPTIDE.  
 FT CHAIN 90 110 INSULIN 1 A CHAIN.  
 FT DISULFID 31 96 INTERCHAIN.  
 FT DISULFID 43 109 INTERCHAIN.  
 FT DISULFID 95 100  
 SQ SEQUENCE 110 AA; 12420 MW; 51D606DA54AE3533 CRC64;

Query Match 83.2%; Score 185; DB 1; Length 110;  
 Best Local Similarity 83.7%; Pred. No. 2e-34;  
 Matches 72; Conservative 4; Mismatches 10; Indels 0; Gaps 0;  
 OY 1 FVNQHIGZGSHLEALYLVCGERGFFYTPKTRKEADLVQGVGGGPGAGSLQPLALEG 60  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 DB 25 FVKQHLGCPHLVEALYLVCGERGFFYTPKSRREVDPQVPLELGGGPGAGDLQTLALEV 84  
 OY 61 SLOKRGIVEQCCTICSLSLYOLENYCN 86  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 DB 85 ARQKRGIVDCCCTICSLSLYOLENYCN 110  
 RESULT 14  
 INS\_PIG  
 ID INS\_PIG STANDARD; PRT: 138 AA.  
 AC P01315; Q9TSJ5;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DI 16-OCT-2001 (Rel. 40, Last sequence update)  
 DI 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Insulin precursor.  
 GN INS.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eularchia; Eupartiodactyla; Suidae; Suidae;  
 OC Sus; Sus; Sus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Han X.G., Tsch B.E.;  
 RT "Complete porcine preproinsulin cDNA sequence.";  
 RL Submitted (MAY 1998) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN: Large white;  
 RX MEDLINE=22135958; PubMed=12140686;  
 RA Amarger V., Nguyen M., Lacroix A.S., Brunschweig M., Nezer C.,  
 RA Georges M., Andersson L.;  
 RT "Comparative sequence analysis of the INS IGF2-H19 gene cluster in  
 RT pigs.";  
 RL Mamm. Genome 13:388-398(2002).  
 RN [3]  
 RP SEQUENCE OF 25-108  
 RX MEDLINE=68286485; PubMed=5657063;  
 RA Chance R.E., Ellis R.M., Bromer W.W.;  
 RT "Porcine preinsulin: characterization and amino acid sequence.";  
 RL Science 161:165-167(1968).  
 RN [4]  
 RP REVISION TO 59.  
 RA Chance R.E.;  
 RL Submitted (JUL-1970) to the PIR data bank.  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
 RA Blundell T.L., Dodson G.G., Hodgkin D., Mercola D.;  
 RT "Insulin. The structure in the crystal and its reflection in  
 RT chemistry and biology.";  
 RL Adv. Protein Chem. 26:279-402(1972).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).  
 RA Isaacs N.W., Agarwal R.C.;  
 RT "Experience with fast Fourier least squares in the refinement of the  
 RT crystal structure of rhombohedral 2-zinc insulin at 1.5-A  
 RT resolution.";  
 RL Acta Crystallogr. A 34:782-791(1978).  
 RN [7]  
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).  
 RX MEDLINE=89099318; PubMed=2905485;  
 RA Baker E.N., Blundell T.L., Cutfield J.F., Cutfield S.M., Dodson E.J.,  
 RA Dodson G.G., Crowfoot Hodgkin D.M., Hubbard R.E., Isaacs N.W.,  
 RA Reynolds C.D., Sakabe K., Sakabe N., Vijayan N.M.;  
 RT "The structure of 2Zn pig insulin crystals at 1.5-A resolution.";  
 RL Philos. Trans. R. Soc. Lond., B, Biol. Sci. 319:369-456(1988).  
 RN [8]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS)



	Matches	70	Conservative	5	Mismatches	11	Indels	0	Gaps	0
Qy	1	FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG	60							
Db	25	FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG	84							
Qy	61	SIQKRGIVEQCCTGICSLYLENYCN	86							
Db	85	ARQKRGIVEQCCTGICSLYLENYCN	111							

Search completed: September 15, 2003, 12:01:40  
Job time : 9.01434 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:54:30 : Search time 33.5986 Seconds  
(without alignments)  
660,520 Million cell updates/sec

Title: US-09-423-100-4

Perfect score: 463

Sequence: 1 FVNHLCGSHLVKALVLCG.....IVQCCTSLCYOLENYCN 86

Scoring table: BL-SUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 260600000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_23:\*

1: sp-archaea:\*

2: sp-bacteria:\*

3: sp-fungi:\*

4: sp-human:\*

5: sp-invertebrate:\*

6: sp-mammal:\*

7: sp-mhc:\*

8: sp-organelle:\*

9: sp-phage:\*

10: sp-plant:\*

11: sp-rodent:\*

12: sp-virus:\*

13: sp-vertebrate:\*

14: sp-unclassified:\*

15: sp-rvirus:\*

16: sp-bacteriap:\*

17: sp-archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY :

Result No.	Score	Match	Length	DB ID	Description
1	463	100.0	110	6	Q8HXV2
2	414	89.2	110	11	Q91X13
3	408	83.8	110	6	Q9NWK6
4	342	73.9	65	6	Q8Z81
5	342	73.9	65	6	Q8Z80
6	246.5	53.2	106	13	Q91R07
7	235.5	50.9	111	13	Q98TA7
8	230.5	49.8	110	13	Q98TA8
9	222.5	48.1	110	13	Q902Y1
10	219	47.3	111	13	Q98T50
11	214.5	46.3	108	13	Q9DD55
12	212.5	45.9	108	13	Q902N4
13	210.5	45.5	87	13	Q98TA9
14	205.5	44.4	108	13	Q98T51
15	203.5	44.0	91	13	Q98T52
16	189	40.8	41	11	Q62543

17	162	35.0	39	11	Q62542
18	142.5	30.8	159	13	Q93607
19	142.5	30.8	182	13	Q73720
20	142.5	30.8	182	13	Q42289
21	142.5	30.8	182	13	P75824
22	142.5	30.8	185	13	O57436
23	142.5	30.8	186	13	Q93527
24	141	30.5	207	13	Q90XD0
25	140.5	30.3	132	13	Q8AV14
26	138.5	29.9	153	13	Q93380
27	137	29.6	185	13	Q9V157
28	137	29.6	210	13	Q91443
29	136.5	29.5	62	13	Q91AA0
30	136.5	29.5	116	13	Q91161
31	136.5	29.5	117	13	Q91476
32	136.5	29.5	145	13	Q91475
33	136.5	29.5	149	13	Q91231
34	136.5	29.5	155	13	Q91162
35	136.5	29.5	161	13	Q91230
36	136.5	29.5	168	13	P81268
37	136.5	29.5	168	13	Q91965
38	135.5	29.3	184	13	Q42336
39	135.5	29.4	184	13	Q9PSX5
40	135	29.2	212	13	Q8JIE4
41	133.5	28.8	106	6	Q9MY26
42	133.5	28.8	117	13	Q91914
43	133.5	28.8	161	13	Q90VV9
44	133.5	28.8	161	13	Q9PWK2
45	133.5	28.8	161	13	Q98SM6

# ALIGNMENTS

RESULT 1

Q8HXV2  
10 Q8HXV2 PRELIMINARY; PRT; 110 AA.  
AC Q8HXV2: 2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, last sequence update)  
UT 01-MAR-2003 (TrEMBLrel. 24, last annotation update)  
DE Insulin precursor.  
GN INS.  
OS Pongo pygmaeus (orangutan).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.  
CX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Stead J.D.H., Jeffreys A.J.;  
RT Haplotype diversity at the insulin region.  
KL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AY37503; AAN06937.1; -;  
SQ SEQUENCE 110 AA; 2202H42H94F520F8 CRC64;

Query Match 100.0%; Score 463; DB 6; Length 110;

Best Local Similarity 100.0%; Pred No 39e-47;

Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVKALVLCGGERGFYTPKTRKREARLQVGVLEGGSPAGSLQPLALEG 60

.....

DB 25 FVNHLCGSHLVKALVLCGGERGFYTPKTRKREARLQVGVLEGGSPAGSLQPLALEG 84

QY 61 SLOKRGIVQCTSLCYOLENYCN 86

.....

DB 85 SLOKRGIVQCTSLCYOLENYCN 110

RESULT 2

Q91X13  
10 Q91X13 PRELIMINARY; PRT; 110 AA.

AC Q91X13:

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

```

DT 01-DEC-2001 (TRENBLREL, 19, Last sequence update)
DE Insulin.
OS Spermophilus tridecemlineatus (Thirteen-lined ground squirrel).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;
OC Spermophilus.
OX NCBI_TaxID=43179;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Tredrea M.M., Buck M.J., Gubanyiogi J., Squire T.L., Andrews M.T.;
RT "Molecular evolution in a hibernating mammal";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
DR EMBL: AY038604; AAK72558.1;
DR HSSP: P01308; INP.
DR InterPro: IPR004875; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR0277; INSULIN.
DR SMART: SM0076; IIGF; 1.
DR PROSITE: PS_0262; INSULIN; 1.
SQ SEQUENCE 110 AA: 12004 MW: 45117.6866422065 CRC64;
      Query Match      89.2%  Score 414; DB 11; Length 110;
      Best Local Similarity 89.5%  Pred. No. 3,60-41;
      Matches 77; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPAGCSLQPLALEG 60
DB 25 FVNHLCGSHLVEALYLVCGERGFFYTPKSPREVEEUGGVELGGPAGCSLQPLALEM 84

QY 61 SLOKRGIVQCCCTSCSLYQLENYCN 86
DB 85 PLOKRGIVEQCASVCSLYQLEHYCN 110

RESULT 4
QY 61 SLOKRGIVQCCCTSCSLYQLENYCN 86
DB 85 PLOKRGIVEQCASVCSLYQLEHYCN 110

PRELIMINARY: PRT; 65 AA.
AC Q8H281;
DT 01-MAR-2003 (TRENBLREL, 23, Created)
DT 01-MAR-2003 (TRENBLREL, 23, Last sequence update)
DE Insulin (Fragment).
OS Gorilla gorilla (gorilla).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Gorilla.
OX NCBI_TaxID=9593;
RN [1]
RP SEQUENCE FROM N.A.
RC Gorilla C. Tiedje H. Klein J.;
RT "Molecular evolution in higher primates: gene specific and organism specific characteristics";
RL Submitted (MAR 2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY092024; AAM76640.1;
DR NON_TER 1;
DR NON_TER 65;
SQ SEQUENCE 65 AA: 6920 MW: 8772017ED8BCA8EA CRC64;
      Query Match      73.9%  Score 342; DB 6; Length 65;
      Best Local Similarity 100.0%  Pred. No. 4,7e-13;
      Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPAGCSLQPLALEGSLKRG 66
DB 1 CGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPAGCSLQPLALEGSLKRG 60

QY 67 IVEQC 71
DB 61 IVEQC 65

RESULT 5
QY 67 IVEQC 71
DB 61 IVEQC 65

PRELIMINARY: PRT; 65 AA.
AC Q8H280;
DT 01-MAR-2003 (TRENBLREL, 23, Created)
DT 01-MAR-2003 (TRENBLREL, 23, Last sequence update)
DE Insulin (Fragment).
OS Pongo pygmaeus (orangutan).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pongo.
OX NCBI_TaxID=9600;
RN [1]
RP SEQUENCE FROM N.A.
RC Gorilla C. Tiedje H. Klein J.;
RT "Molecular evolution in higher primates: gene specific and organism specific characteristics";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY092024; AAM76641.1;
DR NON_TER 1;
DR NON_TER 65;
SQ SEQUENCE 65 AA: 6920 MW: 8772017ED8BCA8EA CRC64;
      Query Match      73.9%  Score 342; DB 6; Length 65;
      Best Local Similarity 100.0%  Pred. No. 4,7e-13;
      Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPAGCSLQPLALEGSLKRG 66
DB 1 CGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPAGCSLQPLALEGSLKRG 60

```



OC Actinopterygii: Neopterygii: Teleostei: Osteoglossomorpha;  
 OC Osteoglossiformes: Hiodontidae; Hiodon.  
 OX NCBI\_TaxID=54904;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21203577; PubMed=11306171;  
 RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;  
 RT \*Molecular cloning of preproinsulin cDNAs from several  
 RT osteoglossomorphs and a cyprinid.  
 RL Mol. Cell. Endocrinol. 174:51-58(2001).  
 CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 DR EMBL: AF282408; AAK34684.1;  
 DR HSSP: P01308; ILPH.  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1.  
 DR PRINTS: PR00277; INSULIN.  
 DR SMART: SM00078; IIGF; 1.  
 DR PROSITE: PS00262; INSULIN; 1.  
 FT NON\_TER 119 110  
 FT SEQUENCE 119 AA: 1244 MW: 106526.956 GRC64;  
 SQ

Query Match 49.1%; Score 242.5; DB 13; Length 111;  
 Best Local Similarity 45.4%; Pred. No. 3;e-18;  
 Matches 44; Conservative 13; Mismatches 15; Indels 25; Gaps 3;

QY 3 NOHLCGSHLVEALYVCGERGFFYTPKTRFAEDLVQVQLGGPGAGSLQPLAL;---- 58  
 DB 26 SOHLGSHLVADALYVCGERGFFYTPKTRFAEDLVQVQLGGPGAGSLQPLAL;---- 58

QY 59 -----PCSLQ-KRGIVQCCSTICSLSYLENYCN 86  
 DB 74 ADYPPYKUGDLKVRGIVECCCHPCNIFDLNVCN 110

# RESULT 10

ID Q98TB0 PRELIMINARY; PRT: 111 AA.  
 AC Q98TB0;  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Preproinsulin (Fragment).  
 OS Chitalla chitula (clown knifefish).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;  
 OX NCBI\_TaxID=12163;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21203577; PubMed=11306171;  
 RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;  
 RT \*Molecular cloning of preproinsulin cDNAs from several  
 RT osteoglossomorphs and a cyprinid.  
 RL Mol. Cell. Endocrinol. 174:51-58(2001).  
 CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 DR EMBL: AF199586; AAK28710.1;  
 DR HSSP: P01308; ILPH.  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1.  
 DR SMART: SM00078; IIGF; 1.  
 FT NON\_TER 111 111  
 FT SEQUENCE 111 AA: 1248 MW: 247644.433 GRC64;  
 SQ

Query Match 47.1%; Score 239; DB 13; Length 111;  
 Best Local Similarity 49.0%; Pred. No. 4;e-18;  
 Matches 48; Conservative 7; Mismatches 17; Indels 26; Gaps 4;

QY 3 NOHLCGSHLVEALYVCGERGFFYTPKTRFAEDLVQVQLGGPGAGSLQPLAL;LEG 60  
 DB 26 NOHLCGSHLVEALYVCGERGFFYTPKTRFAEDLVQVQLGGPGAGSLQPLAL;LEG 60

QY 61 SL-----OKRGIVECCSTICSLSYLENYCN 86  
 DB 74 EVDFYPPKUGDLKVRGIVECCCHPCNIFDLNVCN 111

## RESULT 11

ID Q9DDE5 PRELIMINARY; PRT: 108 AA.  
 AC Q9DDE5;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Insulin precursor.  
 GN INS.  
 OS Brachydanio rerio (zebrafish) (Danio rerio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 OX NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95425199; PubMed=10495243;  
 RA Arntson F., Zechin F., Hortonissi M.;  
 RT \*Early appearance of pancreatic hormone expressing cells in the  
 RT zebrafish embryo.  
 RL Mech. Dev. 87:217-223(1999).  
 CC -1- SURCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 DR EMBL: AJ237750; CAC0109.1;  
 DR HSSP: P01308; ILPH.  
 DR ZFIN: ZDB-GENE-980526-110; ins.  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1.  
 DR PRINTS: PR00277; INSULIN.  
 DR SMART: SM00078; IIGF; 1.  
 DR PROSITE: PS00262; INSULIN; 1.  
 KW Signal.  
 FT SIGNAL 1 23 POTENTIAL  
 FT CHAIN 24 53 INSULIN B CHAIN.  
 FT CHAIN 86 108 INSULIN A CHAIN.  
 SQ SEQUENCE 108 AA: 11904 MW: 319526.957 GRC64;  
 Query Match 46.3%; Score 214.5; DB 13; Length 108;  
 Best Local Similarity 45.8%; Pred. No. 3;e-17;  
 Matches 44; Conservative 11; Mismatches 14; Indels 27; Gaps 3;

QY 4 OHLCGSHLVEALYVCGERGFFYTPKTRFAEDLVQVQLGGPGAGSLQPLAL;EGS-- 61  
 DB 27 OHLCGSHLVADALYVCGERGFFYTPKTRFAEDLVQVQLGGPGAGSLQPLAL;EGS-- 61

QY 62 -----LOKRGIVECCSTICSLSYLENYCN 86  
 DB 73 ADFAFKDHAELIPKRGIVECCCHPCNIFDLNVCN 108

## RESULT 12

ID Q90ZM4 PRELIMINARY; PRT: 108 AA.  
 AC Q90ZM4;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Preproinsulin.  
 OS Catla catla (catla).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Catla.  
 OX NCBI\_TaxID=72446;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX Bhattacharya S., Roy S.S., Dasgupta S., Ravikumar L., Mukherjee M.,  
 RA Bandyopadhyaya L., Wakabayashi K.;  
 RT \*A new cell secreting insulin.\*



Best Local Similarity 46.7%, Pred. No. 1.8e-16;  
Matches 42; Conservative 13; Mismatches 26; Indels 9; Gaps 4;  
QY 4 QHLCGSHLVEALYLVCGERGFFYTPKTAREADLQ---VGQVELGGGPGA-GSLQPLALE 59  
          |||||:|||||:||||| 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42  
Db 4 QHLCGSHLVDALYLVCGRGFFYNPK--RDVDL--GFLPPKADGAAAPGGGENEVAEFAPK 61  
          |||||:|||||:||||| 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42  
QY 60 GSLO---KRGIVEQCCTSICSLYOLENYCN 86  
          :: ||||| 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42  
Db 62 QMEMWVKRGIVEQCCHHPNCNIFDLGRYCN 91  
          :: ||||| 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42

Search completed: September 15, 2003, 12:03:29  
Job time : 45.5986 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:44:15 : Search time 36.5305 Seconds  
(without alignments)  
225.942 Million cell updates/sec

Title: US-09-423-100-5

Perfect score: 294

Sequence: 1 FVNHLCGSHLEALYLWCG.....IVEQCCTSLCYLQLENYCN 52

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107864 seqs, 154726573 residues

Total number of hits satisfying chosen parameters: 1117864

Minimum DB seq length: 9

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*

1: /SIDSL/qcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSL/qcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSL/qcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSL/qcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSL/qcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSL/qcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSL/qcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSL/qcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSL/qcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSL/qcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSL/qcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSL/qcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSL/qcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSL/qcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSL/qcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSL/qcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSL/qcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSL/qcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSL/qcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSL/qcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSL/qcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSL/qcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSL/qcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSL/qcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	294	100.0	52	AA1980	Human insulin prec
2	294	100.0	56	AA1981	Human pro-insulin
3	294	100.0	56	AA1982	Proinsulin sequence
4	294	100.0	63	AA1983	Human pro-insulin
5	294	100.0	96	AA1984	Human pro-insulin
6	294	100.0	96	AA1985	Fusion protein con
7	294	100.0	107	AA1986	hGH-mini-proinsuli
8	294	100.0	116	AA1987	SOD-proinsulin hyb
9	294	100.0	137	AA1988	Mating factor alph

10	294	100.0	145	AA1989	Mating factor alph
11	294	100.0	146	AA1990	Mating factor alph
12	294	100.0	150	AA1991	Chimeric protein,
13	291	99.0	57	AA1992	Proinsulin analog
14	287	97.6	52	AA1993	Example of human i
15	283.5	96.4	53	AA1994	Di-Arg-(R11-32)-Hu
16	283.5	96.4	53	AA1995	Ins double-chain
17	283.5	96.4	53	AA1996	Glycosylphosphatid
18	282.5	96.1	160	AA1997	SOD-proinsulin hyb
19	282.5	95.7	60	AA1998	Human proinsulin a
20	278.5	94.7	51	AA1999	Human mature insul
21	278.5	94.7	51	AA2000	Human mature insul
22	278.5	94.7	51	AA2001	Human insulin, Ho
23	278.5	94.7	53	AA2002	Modified pJB59 mod
24	278.5	94.7	55	AA2003	Modified pJB59 mod
25	278.5	94.7	63	AA2004	Single chain insul
26	278.5	94.7	66	AA2005	Met-Arg-Met-singl
27	278.5	94.7	67	AA2006	Single chain insul
28	278.5	94.7	148	AA2007	pK142 modified in
29	277.5	94.4	63	AA2008	Sequence of novel
30	277.5	94.4	63	AA2009	Gly-A21-di-Arg-(B3
31	277	94.2	65	AA2010	Proinsulin, 1
32	276.5	94.0	67	AA2011	Ins double-chain
33	276.5	94.0	67	AA2012	Ins2 double-chain
34	275.5	93.7	53	AA2013	Example of human i
35	275.5	93.7	53	AA2014	PAK679 modified in
36	275.5	93.7	53	AA2015	Modified human ins
37	275.5	93.7	53	AA2016	Amino acid sequenc
38	275.5	93.7	65	AA2017	Sequence encoded b
39	275.5	93.7	58	AA2018	Modified pJB59 mod
40	275.5	93.7	59	AA2019	Modified pJB59 mod
41	275.5	93.7	61	AA2020	Modified pJB59 mod
42	275.5	93.7	65	AA2021	N-terminally exten
43	275.5	93.7	89	AA2022	Signal peptide/lea
44	275.5	93.7	91	AA2023	Insulin precursor
45	275.5	93.7	91	AA2024	YAP3 signal peptid

ALIGNMENTS

RESULT 1  
AA1980  
ID: AA1980 standard: protein; 52 AA.  
AC: AA1980  
DT: 19-JAN 2000 (first entry)  
DE: Human insulin precursor, SEQ ID 5.  
KW: Insulin; precursor; growth hormone; chaperone; intramolecular;  
KW: folding; conformation; chimeric protein; cleavable; recombinant;  
KW: production; yield.  
OS: Homo sapiens.  
PN: W09950302-A1.  
XX: 07-OCT-1999.  
XX: 31-MAR-1998; 98WO-CN00052.  
XX: 31-MAR-1998; 98WO-CN00052.  
XX: (TONG) TONGHUA GANTECH BIOTECHNOLOGY LTD.  
XX: Gan Z;  
XX: WPI: 1999-610839/52.  
XX: New chimeric proteins containing human growth hormone fragment, used  
PT: particularly for the production of human insulin

fewer necessary procedural steps and hence resulting in a higher yield of human insulin. The INC sequences not only protect insulin sequences from intracellular degradation by a microorganism host, but also promote the folding of the fused insulin precursor, facilitate the solubility of the fusion protein and decrease the intermolecular interactions among the fusion proteins, thus allowing refolding of the fused insulin precursor at comparatively set high concentrations. The procedural steps of: (a) expression in *Escherichia coli*, (b) oxidative solubilization and (c) refolding are the same as those described previously (19). The purification steps can be performed, along with the use of high concentrations of imidazole, in a buffer of high ionic strength, thus not

AAK78665  
ID AAK78665 standard: protein; AA.  
XX

XX  
DT C3-APR-1996 (first entry)  
XX  
XX PROPOSITION sequence 3

KW Proinsulin; post-translational modification; recombinant  
 KW protein folding; conformation  
 XX  
 XX Synthesis

Key	Location/Qualifiers
XX	1..4
FH	/label = k2
FT	/notes = a peptide of 4 amino acids
FT	

FI Peptide  
5.1.34  
Label = P1-(22-238)-Y

FT	Region	35	Diabetes	Diabetes Y	Diabetes Y
FT	Region	35	Diabetes	Diabetes Y	Diabetes Y
FT	Region	35	Diabetes	Diabetes Y	Diabetes Y
FT	Region	35	Diabetes	Diabetes Y	Diabetes Y

36.56  
peptide  
C10H17N3O5

FT yy /note= "human insulin A-chain"

PN EP668292 A2.

PD 23-AUG 1995.

PF 09-FEB-1995; 95EP-0101748.

PR 18-FEB-1994: 941E 405179.

XX  
PA  
/ CARH / HOFERST A2[illegible]XX  
XX

Isolation of insulin that is correctly post-translationally processed - by reacting pro:insulin with a mercaptan in the presence of a chaotropic agent and purification after absorption to hydrophobic ion exchange resin.

pt resin  
yy



Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIICSLYLENYCN 52  
 DB 45 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIICSLYLENYCN 96

RESULT 6  
 AAK78662  
 ID AAK78662 standard; protein: 96 AA.  
 AC AAK78662;  
 XX  
 XX  
 DT 03-APR-1996 (first entry)  
 XX  
 DE Fusion protein contg. proinsulin sequence 3.  
 XX  
 KW Proinsulin; post-translational modification; recombinant production;  
 KW protein folding; conformation;  
 XX  
 OS Synthetic;  
 XX  
 FH Key location/Qualifiers  
 FT Region 41..44  
 FT /label= R2  
 FT /note= "a peptide of 4 amino acids"  
 FT Peptide 45..74  
 FT /label= R1-(R2-R29)-Y  
 FT /note= "human insulin B-chain"  
 FT Region 75  
 FT /label= X  
 FT Peptide 76..96  
 FT /label= Gly (A2-A20)-R4  
 FT /note= "human insulin A-chain"  
 XX  
 PN EP648292-A2;  
 XX  
 XX 23 AUG-1995;  
 XX  
 XX 09-FEB-1995; 95EP-0101748;  
 XX  
 XX 18 FEB-1994; 94DE-4405179;  
 XX (FARH ) H6BCRST AG;  
 PA  
 PI Gerl M, Ludwig J, Obermeier R, Sabat W;  
 XX WP: 1995-284754/38;  
 DR  
 XX  
 XX Isolation of insulin that is correctly post translationally  
 PT processed by reacting proinsulin with a mercaptan in the presence  
 PT of a chaotropic agent and purification after absorption to hydrophobic  
 PT resin  
 XX  
 PS Example 2; Page 8; 16pp; German;  
 CC  
 CC The present sequence is that of a fusion protein, produced in E.coli  
 CC which contains an example of a proinsulin molecule corresp.  
 CC to the general formula R2-R1-(B2-B29) Y-X-Gly-(A2-A20)-R3 (17). In  
 CC formula (17), X = Lys, Arg or a peptide of 2-35 amino acids contg.  
 CC Lys or Arg at the N- and C-termini; Y = a natural amino acid; R1 =  
 CC Phe or a bond; R2 = H, Arg, Lys, a peptide of 2-45 amino acids  
 CC contg. Arg or Lys at the N- and C-termini; R3 = a natural amino  
 CC acid; (A2-A20) and (B2-B29) are the insulin A- and B-chain sequences  
 CC from human or other insulin. The proinsulin molecule, released by  
 CC cyanogen bromide, is reacted with mercaptan at a ratio of 2-10 SH  
 CC residues of mercaptan per Cys residue of proinsulin. The reaction  
 CC takes place in the presence of a chaotropic auxiliary agent at  
 CC pH 10-11 and results in proinsulin with correctly linked cystine  
 CC bridges. Reaction with trypsin and opt. carboxypeptidase B yields  
 CC correctly folded insulin. The insulin is isolated by absorption on  
 CC a hydrophobic resin.

SQ Sequence 96 AA;

Query Match 100.0%; Score 294; DB 16; Length 96;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-26;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIICSLYLENYCN 52  
 DB 45 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIICSLYLENYCN 96

## RESULT 7

AA42860  
 ID AAY42860 standard; protein: 107 AA.

XX  
 AC AAY42860;  
 XX

DT 19 JAN-2000 (first entry)  
 XX

DE hGH mini-proinsulin chimeric protein;  
 XX

KW Insulin precursor; growth hormone; chaperone; intramolecular;  
 KW folding; conformation; chimeric protein; cleaved; recombinant;  
 XX production; yield;  
 XX

OS Synthetic;  
 OS Homo sapiens;  
 XX

PN W09550302-A1;  
 XX

PD 07-04-1999;  
 XX

PF 31-MAR-1998; 98WO 006052;  
 XX

PR 31-MAR 1998; 98WO 006052;  
 XX

PA (TONG-) TONGHUA GENE-TECH BIOTECHNOLOGY LTD.  
 XX

PI Gan Z;  
 XX

DR WPI: 1999-610839/52.  
 XX

PT New chimeric proteins containing human growth hormone fragment, used  
 PT particularly for the production of human insulin;  
 XX  
 PS Claim 13; Page 30; 45pp; English;  
 CC

CC This sequence represents a chimeric protein, hGH-mini-proinsulin.  
 CC This chimeric protein contains an N-terminal fragment of human growth  
 CC hormone (hGH) of the sequence given in AAY42855, a cleavable peptide  
 CC linker (AAY42857), and a human insulin precursor comprising insulin  
 CC A and B chains (AAY42859). The hGH portion of the chimeric protein acts  
 CC as an intramolecular chaperone (IMC) for the insulin precursor,  
 CC enabling it to fold correctly. The cleavable peptide linker has a  
 CC C-terminal Arg residue which enables the hGH portion of the  
 CC chimeric protein to be removed after folding has taken place. Production  
 CC of recombinant human insulin via an hGH proinsulin chimeric protein can  
 CC provide human insulin with correctly linked cysteine bridges with  
 CC fewer necessary procedural steps, and hence resulting in a higher yield  
 CC of human insulin. The IMC sequences not only protect insulin sequences  
 CC from intracellular degradation by a microorganism host, but also promote  
 CC the folding of the fused insulin precursor, facilitate the solubility of  
 CC the fusion protein and decrease the intermolecular interactions among  
 CC the fusion proteins, thus allowing folding of the fused insulin precursor  
 CC at commercially useful high concentrations. The procedural steps of  
 CC cyanogen bromide cleavage, oxidative sulfitolysis and related  
 CC purification steps can thus be eliminated, along with the use of high  
 CC concentrations of mercaptan or the use of hydrophobic absorbent resins.

SQ Sequence 107 AA;

Query Match 100.0%; Score 294; DB 20; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-26;

Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 52  
 DB 56 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 107

RESULT 8  
 AAR98897  
 ID AAR98897 standard: Protein: 116 AA.  
 XX AC AAR98897;  
 XX DT 25-NOV-2003 (updated)  
 XX DT 20-NOV-1995 (first entry)  
 XX DE Mating factor alpha 1-Insulin precursor ArgB31.  
 XX KW Human insulin precursor ArgB31; diabetes; zinc ion complex;  
 XX KW mating factor alpha 1.  
 XX OS Homo sapiens.  
 XX FH Key Location/Qualifiers  
 XX FT Protein 1..85  
 XX FT /label= mating factor alpha 1  
 XX FT /peptide 96..116  
 XX FT /label= B chain  
 XX FT /peptide 117..137  
 XX FT /label A chain  
 XX PN W09507431 A1.  
 XX PD 23 MAR-1995.  
 XX PF 16-SEP-1994; 94WO-060347.  
 XX PR 17-SEP-1993; 93DK-0601044.  
 XX PR 02 FEB-1994; 94OS-0190829.  
 XX PA (NOVO) NOVO-NORDISK AS.  
 XX PI Andersen AS, Walstrom JB, Havelund S, Jonassen L;  
 XX PI Markussen J;  
 XX DR WPI: 1995-131314/17.  
 XX DR N-PSDB: AA086425.  
 XX PT Acylated insulin deriv. which may be present as a zinc ion  
 XX PT complex - is used to treat diabetes and is rapid acting.  
 XX PS Example 14; Figure 7; 69pp; English.  
 XX CC A new method for the production of recombinant human insulin  
 XX CC comprises folding a hybrid polypeptide comprising proinsulin under  
 XX CC conditions that permit correct disulphide bond formation and  
 XX CC subjecting that folded protein to enzymatic cleavage. The insulin  
 XX CC produced can then be purified. This sequence is a SOD-insulin B  
 XX CC chain-Arg-insulin A chain hybrid polypeptide and is encoded by  
 XX CC the plasmid construct pDBAST-LAT. Transformation of the proper  
 XX CC E.coli host cells with pDBAST-LAT results in the efficient  
 XX CC expression of the proinsulin hybrid polypeptide, useful for human  
 XX CC insulin production. The method produces recombinant human insulin  
 XX CC identical to the natural hormone. Hazardous and cumbersome  
 XX CC procedures involving cyanogen bromide and sulphitolysis to protect  
 XX CC SH groups are avoided since the entire hybrid polypeptide folds  
 XX CC efficiently to the native structure even with the leader attached  
 XX CC and Cys unprotected.  
 XX SQ Sequence 116 AA;  
 Query Match 100.0%; Score 293; DB 17; Length 116;  
 Best Local Similarity 100.0%; Pred. No. 3,5e-26;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 52  
 DB 65 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 116

RESULT 9  
 AAR71692  
 ID AAR71692 standard: Protein: 145 AA.  
 XX AC AAR71692;  
 XX DT 25-MAR-2003 (updated)  
 XX DT 20-NOV-1995 (first entry)  
 XX DE Mating factor alpha 1-Insulin precursor ArgB1, ArgB31 N-terminal.

ID AAR71692 standard: Protein: 137 AA.  
 XX AC AAR71692;  
 XX DT 25-NOV-2003 (updated)  
 XX DT 20-NOV-1995 (first entry)  
 XX DE Mating factor alpha 1-Insulin precursor ArgB31.  
 XX KW Human insulin precursor ArgB31; diabetes; zinc ion complex;  
 XX KW mating factor alpha 1.  
 XX OS Homo sapiens.  
 XX FH Key Location/Qualifiers  
 XX FT Protein 1..85  
 XX FT /label= mating factor alpha 1  
 XX FT /peptide 96..116  
 XX FT /label= B chain  
 XX FT /peptide 117..137  
 XX FT /label A chain  
 XX PN W09507431 A1.  
 XX PD 23 MAR-1995.  
 XX PF 16-SEP-1994; 94WO-060347.  
 XX PR 17-SEP-1993; 93DK-0601044.  
 XX PR 02 FEB-1994; 94OS-0190829.  
 XX PA (NOVO) NOVO-NORDISK AS.  
 XX PI Andersen AS, Walstrom JB, Havelund S, Jonassen L;  
 XX PI Markussen J;  
 XX DR WPI: 1995-131314/17.  
 XX DR N-PSDB: AA086425.  
 XX PT Acylated insulin deriv. which may be present as a zinc ion  
 XX PT complex - is used to treat diabetes and is rapid acting.  
 XX PS Example 5; Page 78; 160pp; English.  
 XX CC AA086425 encodes AAR71692 mating factor alpha 1-Insulin precursor  
 XX CC ArgB31. ArgB31 comprises the B and A chains of a claimed human  
 XX CC insulin derivative. In the final claimed compsn. they are covalently  
 XX CC connected via disulphide bonds between Cys residues A7/B7 and  
 XX CC A20/B19. The derivative, which may be present as a zinc ion  
 XX CC complex, can be used as a fast action treatment for diabetes.  
 XX CC (Updated on 25-MAR-2003 to correct PN field.)  
 XX SQ Sequence 137 AA;  
 Query Match 100.0%; Score 294; DB 16; Length 137;  
 Best Local Similarity 100.0%; Pred. No. 4.1e-26;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 52  
 DB 96 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 137

RESULT 10  
 AAR71694  
 ID AAR71694 standard: Protein: 145 AA.  
 XX AC AAR71694;  
 XX DT 25-MAR-2003 (updated)  
 XX DT 20-NOV-1995 (first entry)  
 XX DE Mating factor alpha 1-Insulin precursor ArgB1, ArgB31 N-terminal.



XX	07 OCT-1999.
PD	
XX	
PF	A1 MAR-1998; 98WO-CN00052.
XX	
PR	C1-MAR-1998; 98WO-CN00052.
XX	
PA	(TONG) TONGHUA CANTECH BIOTECHNOL-XY LTD.
PI	Gan Z;
XX	
DR	WPI: 1999-610839/52.
XX	
PT	New chimeric proteins containing human growth hormone fragment, used
PI	particularly for the production of human insulin
XX	
PS	Claim 14; Page 30-31; 46pp; English.
XX	
CC	This sequence represents a chimeric protein, which contains an
CC	N-terminal fragment of human growth hormone (hGH) of the sequence given
CC	in AAY42856, a cleavable peptide linker (AAY42857), and a human insulin
CC	precursor consisting insulin A and B chains (AAY42858). The B-chain portion
CC	of the chimeric protein acts as an insulin receptor substrate (IRS) for
CC	the insulin precursor, enabling it to bind correctly. The cleavable
CC	peptide linker has a C-terminal Ala residue which enables the B-chain portion
CC	of the chimeric protein to be removed after folding has taken place.
CC	Production of recombinant human insulin via an hGH-proinsulin chimeric
CC	protein can provide human insulin with correctly linked cysteine bridges
CC	with fewer necessary procedural steps, and hence resulting in a higher
CC	yield of human insulin. The hGH sequences not only protect insulin
CC	sequences from intracellular degradation by a microorganism host, but
CC	also promote the folding of the fused insulin precursor, facilitate the
CC	solubility of the fusion protein and decrease the intermolecular
CC	interactions among the fusion proteins, thus allowing folding of the
CC	fused insulin precursor at commercially useful high concentrations. The
CC	procedural steps of cyanogen bromide cleavage, oxidative sulphatolysis
CC	and related purification steps can thus be eliminated, along with the use
CC	of high concentrations of mercaptan or the use of hydrophobic absorbent
XX	resins.
XX	
SQ	Sequence 150 AA:
	Query Match 100.0%; Score 294; DB 20; Length 150;
	Best Local Similarity 100.0%; Pred No 4, 5e-26;
	Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 FVNHLCGSHLEVALYVCGGEGFFYPKTRGRIGVECCCTCSILXOLENVCN 52
	'
Db	99 FVNHLCGSHLEVALYVCGGEGFFYPKTRGRIGVECCCTCSILXOLENVCN 150
RESULT 14	
AA004562	
ID	AA004562 standard; protein; 57 AA.
XX	
AC	AA004582;
XX	
DT	25-MAR-2003 (updated)
DT	14 SEP 1990 (first entry)
XX	
DE	Proinsulin analogue with a Lys residue linking the A and B chains.
XX	
KW	insulin fusion protein; pro-insulin analogue; tandemistate;
KW	Lys Lys bridge, ds
XX	
OS	synthetic.
XX	
XX	Key Location/Qualifiers
PH	misc-difference 16
FT	/label-Lys residue linking insulin B chain to A chain
FT	Peptide 1..35
FT	/label- Insulin B chain
FT	Peptide 37..57



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:57:40 : Search time 11,9283 Seconds  
( without alignments)  
184.449 Million cell updates/sec

Title: US-09-423-100-5

Perfect score: 294

Sequence: 1 FVNHLCGSHLVEALYLVCG.....IVEQCCTICSLYLENYCN 52

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 429717 seqs, 4241658 residues

Total number of hits satisfying chosen parameters 12477

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA\*  
1: /cqn2\_6/plodata/1/iaa/5A... .pep:\*  
2: /cqn2\_6/plodata/1/iaa/5B... .pep:\*  
3: /cqn2\_6/plodata/1/iaa/6A... .pep:\*  
4: /cqn2\_6/plodata/1/iaa/6B... .pep:\*  
5: /cqn2\_6/plodata/1/iaa/PC10S... .pep:\*  
6: /cqn2\_6/plodata/1/iaa/backfiles... .pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	294	100.0	56	1	US-08-160-376A-7 Sequence 7, Appl
2	294	100.0	56	1	US-08-189-487-11 Sequence 11, Appl
3	294	100.0	63	1	US-08-160-376A-6 Sequence 6, Appl
4	294	100.0	66	1	US-08-291-060B-5 Sequence 5, Appl
5	294	100.0	96	1	US-08-160-376A-5 Sequence 5, Appl
6	294	100.0	96	1	US-08-184-487-8 Sequence 8, Appl
7	294	100.0	137	1	US-08-400-256-39 Sequence 39, Appl
8	294	100.0	137	3	US-08-975-365-39 Sequence 39, Appl
9	294	100.0	145	1	US-08-400-256-45 Sequence 45, Appl
10	294	100.0	145	3	US-08-975-365-45 Sequence 45, Appl
11	294	100.0	146	1	US-08-400-256-48 Sequence 48, Appl
12	294	100.0	146	3	US-08-975-365-48 Sequence 48, Appl
13	291	99.0	57	1	US-08-030-741A-44 Sequence 44, Appl
14	283.5	96.4	53	1	US-08-233-617-4 Sequence 4, Appl
15	283.5	96.4	53	4	US-08-981-988A-42 Sequence 42, Appl
16	278.5	94.7	51	4	US-09-472-948-3 Sequence 3, Appl
17	278.5	94.7	51	4	US-09-723-981-3 Sequence 3, Appl
18	277.5	94.7	51	4	US-09-723-896-3 Sequence 3, Appl
19	277.5	94.4	53	1	US-08-233-617-3 Sequence 3, Appl
20	277.5	94.2	65	3	US-08-900-574-3 Sequence 3, Appl
21	276.5	94.0	55	3	US-08-500-574-6 Sequence 6, Appl
22	276.5	94.0	66	3	US-08-900-574-5 Sequence 5, Appl
23	276.5	94.0	67	4	US-08-981-988A-1 Sequence 1, Appl
24	276.5	94.0	67	4	US-08-981-988A-5 Sequence 5, Appl
25	276.5	93.9	67	3	US-08-900-574-7 Sequence 7, Appl
26	275.5	93.7	53	3	US-09-261-853-2 Sequence 2, Appl
27	275.5	93.7	65	1	US-08-468-674B-71 Sequence 71, Appl

Sequence 71, Appl  
Sequence 41, Appl  
Sequence 41, Appl  
Sequence 45, Appl  
Sequence 15, Appl  
Sequence 15, Appl  
Sequence 37, Appl  
Sequence 3, Appl  
Sequence 36, Appl  
Sequence 19, Appl  
Sequence 33, Appl  
Sequence 42, Appl  
Sequence 33, Appl  
Sequence 42, Appl  
Sequence 2, Appl  
Sequence 4, Appl

## ALIGNMENTS

RESULT :  
US-08-160-376A-7  
: Sequence 7, Application US/08160376A  
: Patent No. 5473049  
: GENERAL INFORMATION:  
: APPLICANT: Obermeier, Rancier  
: APPLICANT: Gerl, Martin  
: APPLICANT: Ludwig, Jurgen  
: APPLICANT: Sabel, Walter  
: TITLE OF INVENTION: Process For Obtaining Proinsulin  
: TITLE OF INVENTION: Possessing Correctly Linked  
: TITLE OF INVENTION: Cystine Bridges  
: NUMBER OF SEQUENCES: 7  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Kenneth A. Genoni, Esq.  
: STREET: Rt. 202-06 No. 5473049th/P.O. Box 2500  
: CITY: Somerville  
: STATE: New Jersey  
: COUNTRY: U.S.A.  
: ZIP: 08876-1258  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE  
: COMPUTER: IBM 386  
: OPERATING SYSTEM: WINDOWS 3.1  
: SOFTWARE: WORDPERFECT 5.1  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/160-376A  
: FILING DATE: December 1, 1994  
: CLASSIFICATION: C10  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: GE P 4240420-7  
: FILING DATE: December 2, 1992  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Barbara V. Maurer, Esq.  
: REGISTRATION NUMBER: 31,287  
: REFERENCE/DOCKET NUMBER: HOE 92/F 384  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (908) 231-4079  
: TELEFAX: (908) 231-2255  
: INFORMATION FOR SEQ ID NO: 7:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 56 Amino Acids  
: TYPE: Amino Acid (AA)  
: TOPOLOGY: not relevant  
US-08-160-376A-7

Query Match 100.0% Score 244: DB 1: Length 56:  
Best Local Similarity 100.0% Pred. No. 4.9e-29;  
Matches 52: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

OY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCTICSILYOLENYCN 52  
DB 5 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCTICSILYOLENYCN 56

## RESULT 2

US-08-389-487-11  
Sequence 11, Application US/08389487  
Patent No. 5663291  
GENERAL INFORMATION:  
APPLICANT: Obermeyer, Rainer  
APPLICANT: Gerl, Martin  
APPLICANT: Ludwig, Jurgen  
APPLICANT: Sabel, Walter  
TITLE OF INVENTION: Process for Obtaining Insulin Having  
TITLE OF INVENTION: Correctly Linked Cysteine Bridges  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESSEE: Dunner, L.L.P.  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States of America  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/389,487  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Einaudi, Carol P.  
REGISTRATION NUMBER: 32,220  
REFERENCE/POCKET NUMBER: 02481.1424-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 56 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
MOLECULE TYPE: peptide  
US-08-389-487 11

Query Match 100.0% Score 244, DB 1, Length 56;  
Best Local Similarity 100.0% Pred. No. 4.9e-25;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCTICSILYOLENYCN 52  
DB 5 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCTICSILYOLENYCN 56

## RESULT 3

US-08-160-376A-6  
Sequence 6, Application US/08160376A  
Patent No. 5473049  
GENERAL INFORMATION:  
APPLICANT: Obermeyer, Rainer  
APPLICANT: Gerl, Martin  
APPLICANT: Ludwig, Jurgen  
APPLICANT: Sabel, Walter  
TITLE OF INVENTION: Process For Obtaining Proinsulin  
TITLE OF INVENTION: Possessing Correctly Linked  
TITLE OF INVENTION: Cysteine Bridges  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Kenneth A. Genoni, Esq.  
STREET: Rt. 202-206 No. 5473049th/P.O. Box 2500  
CITY: Somerville  
STATE: New Jersey  
COUNTRY: U.S.A.  
ZIP: 08876-1268  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
COMPUTER: IBM 386  
OPERATING SYSTEM: WINDOWS 3.1  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/160,376A  
FILING DATE: December 1, 1993  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GE P 4240420.7  
FILING DATE: December 2, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Barbara V. Master, Esq.  
REGISTRATION NUMBER: 31,287  
REFERENCE/POCKET NUMBER: HOR 92/F 44  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908) 231-4079  
TELEFAX: (908) 231-2255  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 63 Amino Acids  
TYPE: Amino Acid (AA)  
TOPLOGY: not relevant  
US-08-160-376A-6

Query Match 100.0% Score 244, DB 1, Length 63;  
Best Local Similarity 100.0% Pred. No. 5.6e-25;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCTICSILYOLENYCN 52  
DB 12 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCTICSILYOLENYCN 63

## RESULT 4

US-08-291-060B-5  
Sequence 5, Application US/08291060B  
Patent No. 5728543  
GENERAL INFORMATION:  
APPLICANT: Dorschug, Michael  
APPLICANT: Koller, Klaus-Peter  
APPLICANT: Marquardt, Rudiger  
APPLICANT: Melwes, Johannes  
TITLE OF INVENTION: An Enzymatic Process for the  
TITLE OF INVENTION: Conversion of Preproinsulins Into Insulins  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finneqar, Henderson, Farabow, Garrett &  
ADDRESSEE: Dunner, L.L.P.  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/291,060B  
FILING DATE: 08-AUG-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Einaudi, Carol P.  
REGISTRATION NUMBER: 32,220

```

? REFERENCE/DOCKET NUMBER: 02481.1105-0-0000
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202) 408-4366
? TELEFAX: (202) 408-4400
? INFORMATION FOR SEQ ID NO: 5:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 66 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: peptide
? US-08-291-0608-5

Query Match 100.0%; Score 294; DB 1; Length 66;
Best Local Similarity 100.0%; Pred. No. 5.9e-29;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 52
DB 15 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 66

RESULT 5
US-08-160-376A-5
? Sequence 5, Application US/08160376A
? Patent No. 5471049
? GENERAL INFORMATION:
? APPLICANT: Obermeier, Rainer
? APPLICANT: Gerl, Martin
? APPLICANT: Ludwig, Jurgen
? APPLICANT: Sabel, Walter
? TITLE OF INVENTION: Process for obtaining Proinsulin
? TITLE OF INVENTION: Possessing Correctly Linked
? NUMBER OF SEQUENCES: 7
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Kenneth A. Genoni, Esq
? STREET: Rt. 202-206 No. 5473049H/P.O. Box 2500
? CITY: Somerville
? STATE: New Jersey
? COUNTRY: U.S.A.
? ZIP: 08876-1258
? COMPUTER READABLE FORM:
? MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
? COMPUTER: IBM 386
? OPERATING SYSTEM: WINDOWS 3.1
? SOFTWARE: WORDPERFECT 5.1
? CURRENT APPLICATION DATA:
? FILING DATE: December 1, 1993
? CLASSIFICATION: 530
? PRIOR APPLICATION NUMBER:
? APPLICATION NUMBER: GE P 4240420.7
? FILING DATE: December 2, 1992
? ATTORNEY/AGENT INFORMATION:
? NAME: Barbara V. Maurer, Esq
? REGISTRATION NUMBER: 31,287
? REFERENCE/DOCKET NUMBER: HSE 92/P 364
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (908) 231-4079
? TELEFAX: (908) 231-2255
? INFORMATION FOR SEQ ID NO: 5:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 66 Amino Acids
? TYPE: Amino Acid (AA)
? TOPOLOGY: not relevant
? US-08-160-376A-5

Query Match 100.0%; Score 294; DB 1; Length 96;
Best Local Similarity 100.0%; Pred. No. 4.8e-29;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 52
DB 15 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 96

RESULT 6
US-08-389-487-8
? Sequence 8, Application US/08389487
? Patent No. 5663291
? GENERAL INFORMATION:
? APPLICANT: Obermeier, Rainer
? APPLICANT: Gerl, Martin
? APPLICANT: Ludwig, Jurgen
? APPLICANT: Sabel, Walter
? TITLE OF INVENTION: Process for obtaining Insulin Having
? TITLE OF INVENTION: Correctly Linked Cysteine Bridges
? NUMBER OF SEQUENCES: 12
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
? ADDRESSEE: Dunner
? STREET: 1400 I Street, N.W.
? CITY: Washington
? STATE: D.C.
? COUNTRY: United States of America
? ZIP: 20005-3415
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/189,487
? FILING DATE:
? CLASSIFICATION: 530
? ATTORNEY/AGENT INFORMATION:
? NAME: Elnaudi, Carol P.
? REGISTRATION NUMBER: 32,119
? REFERENCE/DOCKET NUMBER: 12481.1424-000000
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 202-408-4000
? TELEFAX: 202-408-4400
? INFORMATION FOR SEQ ID NO: 8:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 96 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: peptide
? US-08-389-487-8

Query Match 100.0%; Score 294; DB 1; Length 96;
Best Local Similarity 100.0%; Pred. No. 8.8e-29;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 52
DB 45 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 96

RESULT 7
US-08-400-256-39
? Sequence 19, Application US/08400256
? Patent No. 5750497
? GENERAL INFORMATION:
? APPLICANT: Havelund, Svend
? APPLICANT: Halstrom, John
? APPLICANT: Jonassen, Ib
? APPLICANT: Andersen, Asger Sloth
? APPLICANT: Markussen, Jan
? TITLE OF INVENTION: ACYLATED INSULIN
? NUMBER OF SEQUENCES: 49
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: No. 5750497a No. 5750497a disk of No. 5750497th America, Inc.
? STREET: 405 Lexington Avenue, 64th Floor
```

CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/400,256  
FILING DATE: 03-MAR-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 3985, 220-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 45  
SEQUENCE CHARACTERISTICS:  
LENGTH: 145 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-400-256-39

Query Match 100.0% Score 254; DB 1; Length 147;  
Best Local Similarity 100.0%; Pred. No. 1 30-28;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVGFCCTSIQSLYLENYCN 52  
DB 86 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRG VEQCTSIQSLYLENYCN 137

RESULT 8  
US-08-975-365-39  
Sequence 39, Application US/08/975365  
Patent No. 6011007  
GENERAL INFORMATION:  
APPLICANT: Havelund, Svend  
APPLICANT: Halstrom, John  
APPLICANT: Jonassen, Ib  
APPLICANT: Andersen, Asger Sloth  
APPLICANT: Markussen, Jan  
TITLE OF INVENTION: ACYLATED INSULIN  
NUMBER OF SEQUENCES: 49  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6011007o No. 6011007disk of No. 6011007th America, Inc.  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/975,365  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/400,256  
FILING DATE: 03-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 3985, 220-US  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 39;  
SEQUENCE CHARACTERISTICS:  
LENGTH: 137 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-975-365-39

Query Match 100.0% Score 294; DB 4; Length 137;  
Best Local Similarity 100.0%; Pred. No. 1 30-28;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVGFCCTSIQSLYLENYCN 52  
DB 86 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVGFCCTSIQSLYLENYCN 137

RESULT 9  
US-08-400-256-45  
Sequence 45, Application US/08400256  
Patent No. 5750497  
GENERAL INFORMATION:  
APPLICANT: Havelund, Svend  
APPLICANT: Halstrom, John  
APPLICANT: Jonassen, Ib  
APPLICANT: Andersen, Asger Sloth  
APPLICANT: Markussen, Jan  
TITLE OF INVENTION: ACYLATED INSULIN  
NUMBER OF SEQUENCES: 49  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 5750497o No. 5750497disk of No. 5750497th America, Inc.  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/400,256  
FILING DATE: 03-MAR-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 3985, 220-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 45;  
SEQUENCE CHARACTERISTICS:  
LENGTH: 145 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-400-256-45

Query Match 100.0% Score 294; DB 1; Length 145;  
Best Local Similarity 100.0%; Pred. No. 1 40-28;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVGFCCTSIQSLYLENYCN 52  
DB 94 FVNOHLCGSHLVEALYLVCG :RGFFYTPKTRGIVGFCCTSIQSLYLENYCN 145

RESULT 10  
US-08-975-365-45

```

1 MEDIUM TYPE: Floppy disk
2 COMPUTER: IBM PC Compatible
3 OPERATING SYSTEM: PC-DOS/MS-DOS
4 SOFTWARE: PatentIn Release #1.0, Version #1.25
5 CURRENT APPLICATION DATA:
6 APPLICATION NUMBER: US/08/400,256
7 FILING DATE: 03-MAR-1995
8 CLASSIFICATION: 514
9 ATTORNEY/AGENT INFORMATION:
10 NAME: Lambiris, Elias J.
11 REGISTRATION NUMBER: 33,728
12 REFERENCE/DOCKET NUMBER: 3985,220-US
13 TELECOMMUNICATION INFORMATION:
14 TELEPHONE: 212-867-0123
15 TELEFAX: 212-878-9655
16 INFORMATION FOR SEQ ID NO: 48:
17 SEQUENCE CHARACTERISTICS:
18 LENGTH: 146 amino acids
19 TYPE: amino acid
20 TOPLOGY: linear
21 MATERIAL TYPE: protein
22 US 08-400,256-48
23
24 Query Match 100.0% Score 294; DB 1; Length 146;
25 Best Local Similarity 100.0%; Pred. No. 146-28;
26 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
27
28 QY 1 FVNHQAGSHSVVEALYVCGHGFYFPTKRGIVECVTTSISLYOLENYCN 52
29 |||||.....|.....|.....|.....|.....|.....|.....|
30 DB 95 FVNHQAGSHSVVEALYVCGHGFYFPTKRGIVECVTTSISLYOLENYCN 146
31
32 RESULT 12
33 US-08 975-365-48
34 Sequence 48, Application US/08975345
35 Patent No. 6011007
36 GENERAL INFORMATION:
37 APPLICANT: Havelund, Svend
38 APPLICANT: Halstrom, John
39 APPLICANT: Jonassen, Ib
40 APPLICANT: Andersen, Assar Sloth
41 APPLICANT: Markussen, Jan
42 TITLE OF INVENTION: ACRYLATED INSULIN
43 NUMBER OF SEQUENCES: 49
44 CORRESPONDENCE ADDRESS:
45 ADDRESS: No. 6011007o No. 6011007disk of No. 6011007th America, Inc.
46 STREET: 405 Lexington Avenue, 54th Floor
47 CITY: New York
48 STATE: New York
49 COUNTRY: United States of America
50 ZIP: 10174-6401
51 COMPUTER READABLE FORM:
52 MEDIUM TYPE: Floppy disk
53 COMPUTER: IBM PC compatible,
54 OPERATING SYSTEM: PC-DOS/MS-DOS
55 SOFTWARE: PatentIn Release #1.0, Version #1.25
56 CURRENT APPLICATION DATA:
57 APPLICATION NUMBER: US/08/975,365
58 FILING DATE:
59 CLASSIFICATION: 514
60 PRIOR APPLICATION DATA:
61 PRIOR APPLICATION NUMBER: US 08/400,256
62 FILING DATE: 03-MAR-1995
63 ATTORNEY/AGENT INFORMATION:
64 NAME: Lambiris, Elias J.
65 REGISTRATION NUMBER: 33,728
66 REFERENCE/DOCKET NUMBER: 3985,220-US
67 TELECOMMUNICATION INFORMATION:
68 TELEPHONE: 212-867-0123
69 TELEFAX: 212-878-9655
70 INFORMATION FOR SEQ ID NO: 48:
71 SEQUENCE CHARACTERISTICS:
72 LENGTH: 146 amino acids

```

? TYPE: amino acid  
? TOPOLOGY: linear  
? MOLECULE TYPE: protein  
US-08-975-365-48

Query Match 100.0%; Score 294; DB 3; Length 146;  
Best Local Similarity 100.0%; Pred. No. 1.4e-28;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FVNOHLCGSHLVGCGRGGFFYPKTRGIVEOCCTCSICSLYLENYCN 52  
DB 95 FVNOHLCGSHLVGCGRGGFFYPKTRGIVEOCCTCSICSLYLENYCN 145  
|||||

RESULT 13  
US-08-030-731A-44  
? Sequence 44, Application US/08010731A  
? Patent No. 5426046  
? GENERAL INFORMATION:  
? APPLICANT: Kollon, Klaus Peter  
? APPLICANT: Kress, Guenter Hubertus  
? APPLICANT: Krimm, Eugen  
? APPLICANT: Wallmeier, Wolfram  
? TITLE OF INVENTION: Processes for the Preparation of Foreign  
? TITLE OF INVENTION: Proteins in Streptomyces  
? NUMBER OF SEQUENCES: 48  
? CORRESPONDENCE ADDRESS:  
? ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
? ADDRESSEE: Dunner  
? STREET: 1300 I Street, N.W., Suite 700  
? CITY: Washington  
? STATE: D.C.  
? COUNTRY: USA  
? ZIP: 20005-3315  
? COMPUTER READABLE FORM:  
? MEDIUM TYPE: Floppy disk  
? COMPUTER: IBM PC compatible  
? OPERATING SYSTEM: PC-DOS/MS-DOS  
? SOFTWARE: Patent in Release #1.0, Version #1.25  
? CURRENT APPLICATION DATA:  
? APPLICATION NUMBER: US/08/040,741A  
? FILING DATE: 12-MAR-1993  
? CLASSIFICATION: 435  
? PRIOR APPLICATION DATA:  
? APPLICATION NUMBER: US/07/189,840  
? FILING DATE: 03-MAY-1988  
? PRIOR APPLICATION DATA:  
? APPLICATION NUMBER: US/07/430,622  
? FILING DATE: 01-NOV-1989  
? PRIOR APPLICATION DATA:  
? APPLICATION NUMBER: US/07/687,610  
? FILING DATE: 19-APR-1991  
? PRIOR APPLICATION DATA:  
? APPLICATION NUMBER: US/07/735,767  
? FILING DATE: 29-JUL-1991  
? PRIOR APPLICATION DATA:  
? APPLICATION NUMBER: DE P 37 14 866.4  
? FILING DATE: 05-MAY-1987  
? PRIOR APPLICATION DATA:  
? APPLICATION NUMBER: DE P 48 37 273.8  
? FILING DATE: 03-NOV-1988  
? PRIOR APPLICATION DATA:  
? APPLICATION NUMBER: DE P 49 27 445.7  
? FILING DATE: 19-AUG-1989  
? PRIOR APPLICATION DATA:  
? APPLICATION NUMBER: DE P 40 12 816.0  
? FILING DATE: 21-APR-1990  
? ATTORNEY/AGENT INFORMATION:  
? NAME: Kirschner Michael K.  
? REGISTRATION NUMBER: 34,851  
? REFERENCE/DOCKET NUMBER: 02481-0593-02000  
? TELECOMMUNICATION INFORMATION:  
? TELEPHONE: 202-408-4000

? TELEFAX: 202-408-4400  
? INFORMATION FOR SEQ ID NO: 44:  
? SEQUENCE CHARACTERISTICS:  
? LENGTH: 57 amino acids  
? TYPE: amino acid  
? TOPOLOGY: unknown  
? MOLECULE TYPE: peptide  
US-08-030-731A-44

Query Match 99.0%; Score 291; DB 1; Length 57;  
Best Local Similarity 98.1%; Pred. No. 1.2e-28;  
Matches 51; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FVNOHLCGSHLVGCGRGGFFYPKTRGIVEOCCTCSICSLYLENYCN 52  
DB 6 FVNOHLCGSHLVGCGRGGFFYPKTRGIVEOCCTCSICSLYLENYCN 57  
|||||

RESULT 14  
US-08-233-617-4  
? Sequence 1, Application US/08233617  
? Patent No. 5466566  
? GENERAL INFORMATION:  
? APPLICANT: Obermeyer, Rainer  
? APPLICANT: Sabel, Walter  
? APPLICANT: Deil, Peter  
? APPLICANT: Geisen, Ralf  
? TITLE OF INVENTION: Amphiphous Microspherical Forms of Insulin  
? TITLE OF INVENTION: Derivatives  
? NUMBER OF SEQUENCES: 4  
? CORRESPONDENCE ADDRESS:  
? ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
? ADDRESSEE: Dunner  
? STREET: 1300 I Street, N.W., Suite 700  
? CITY: Washington  
? STATE: D.C.  
? COUNTRY: USA  
? ZIP: 20005-3315  
? COMPUTER READABLE FORM:  
? MEDIUM TYPE: Floppy disk  
? COMPUTER: IBM PC compatible  
? OPERATING SYSTEM: PC-DOS/MS-DOS  
? SOFTWARE: Patent in Release #1.0, Version #1.25  
? CURRENT APPLICATION DATA:  
? APPLICATION NUMBER: US/08/233,617  
? FILING DATE: 25-APR-1994  
? CLASSIFICATION: 514  
? PRIOR APPLICATION DATA:  
? APPLICATION NUMBER: P 43 13 752.4  
? FILING DATE: 27-APR-1993  
? ATTORNEY/AGENT INFORMATION:  
? NAME: Carol P. Einaudi  
? REGISTRATION NUMBER: 32,220  
? REFERENCE/DOCKET NUMBER: 02481-1374-00000  
? TELECOMMUNICATION INFORMATION:  
? TELEPHONE: 202 408-4000  
? TELEFAX: 202-408-4400  
? INFORMATION FOR SEQ ID NO: 4:  
? SEQUENCE CHARACTERISTICS:  
? LENGTH: 53 amino acids  
? TYPE: amino acid  
? TOPOLOGY: linear  
? MOLECULE TYPE: protein  
? ORIGINAL SOURCE:  
? ORGANISM: Escherichia coli  
US-08-233-617-4

Query Match 96.4%; Score 283.5; DB 1; Length 53;  
Best Local Similarity 98.1%; Pred. No. 8.9e-28;  
Matches 52; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 FVNOHLCGSHLVGCGRGGFFYPKTRGIVEOCCTCSICSLYLENYCN 52  
|||||

1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRRGIVEQCCTSI CSLYQI.ENYCN 53  
Db

RESULT 15

US-08-981-988A-42

; Sequence 42, Application US/08981988A

; Patent NO. 6337194

; GENERAL INFORMATION:

APPLICANT: Vittal Mallya Scientific Research Foundation

APPLICANT: The University of Leicester

; TITLE OF INVENTION: Insulin

; NUMBER OF SEQUENCES: 43

CORRESPONDENCE ADDRESS:

ADDRESSEE: VITTAL MALLYA SCIENTIFIC RESEARCH FOUNDATION

STREET: K. R. ROAD

CITY: BANGALORE

COUNTRY: INDIA

2. P: 56.0 0.04

# COMPUTER RESEARCH FORM:

MEMORANDUM FOR THE RECORD

COMPL: ER: IM PC: COMPL: FILE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0. Version #1.30 (EPN)

; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/981,988A

FILING DATE:

CLASSIFICATION: 435

PRINTER APPLICATION DATA

APPLICATION NUMBER: CR 9513967.1

FILING DATE: 08-JUL-1995

; INFORMATION FOR SEQ ID NO: 42:

SEQUENCE CHARACTERISTICS:

LENG<sup>3</sup>H: 53 amino acids

LEUO.N: is amino acids  
TYPE: amino acid

TYPE: **STRANDEDNESS:**

STRANDEDNESS: unknown  
TOPOLOGY: unknown

US-08-481-488A-42

### Query Match

Query Match	Score	DB 4: Length
96.48	283	53

Age	Gender	Score	Pred. No.	Score	Pred. No.
18-24	Male	98.1%	98.1%	98.1%	98.1%
18-24	Female	98.1%	98.1%	98.1%	98.1%
25-34	Male	98.1%	98.1%	98.1%	98.1%
25-34	Female	98.1%	98.1%	98.1%	98.1%
35-44	Male	98.1%	98.1%	98.1%	98.1%
35-44	Female	98.1%	98.1%	98.1%	98.1%
45-54	Male	98.1%	98.1%	98.1%	98.1%
45-54	Female	98.1%	98.1%	98.1%	98.1%
55-64	Male	98.1%	98.1%	98.1%	98.1%
55-64	Female	98.1%	98.1%	98.1%	98.1%
65+	Male	98.1%	98.1%	98.1%	98.1%
65+	Female	98.1%	98.1%	98.1%	98.1%

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKT-RGIVECCCTSCSIYOLENYCN 52

QY I FVNQALCGSALVERAIV LVCGENGF IIPKI -RGIVEQGISCSIIQLENIQ

Db 1 FVNQHLCCSHLVEALYLVCGERGFFYTPKT<sup>1</sup>RGIVEQCCTSI<sup>2</sup>CSLYQLENYCN<sup>53</sup>

Search completed: September 15, 2003, 12:05:32

Job time : 11.9283 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: September 15, 2003, 12:03:35 ; Search time 21.8065 Seconds  
(without alignments)  
347.945 Million cell updates/sec

Title: US-09-423-100-5

Perfect score: 294

Sequence: 1 FVNHQHGSHLVEALYLCG.....IVECCSTICSLYOLENYCN 52

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 541946 seqs, 14591242 residues

Total number of hits satisfying chosen parameters: 541946

Minimum DB seq length: 6

Maximum hit seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCI\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/CTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARY

Result No.	Score	Query Match	Length DB	ID	Description
1	294	100.0	52	14	US-10-054-873-5
2	294	100.0	107	14	US-10-054-873-6
3	294	100.0	150	14	US-10-054-873-7
4	278.5	94.7	51	11	US-09-878-978-5
5	278.5	94.7	51	14	US-10-038-473-1
6	275.5	93.7	124	10	US-09-894-711-18
7	275.5	93.7	138	10	US-09-861-687-19
8	273	92.9	50	14	US-10-066-0094-3
9	267	90.8	86	10	US-09-878-380-1
10	267	90.8	86	11	US-09-878-975B-4
11	267	90.8	86	14	US-10-028-410-2
12	267	90.8	86	14	US-10-054-873-4
13	267	90.8	96	10	US-09-947-563-4
14	267	90.8	110	9	US-09-205-658-125
15	267	90.8	110	9	US-09-815-229-3

16	267	90.8	110	10	US-09-804-409A-9	Sequence 9, Appli
17	267	90.8	110	12	US-09-969-748C-6	Sequence 6, Appli
18	267	90.8	110	15	US-10-038-686-1	Sequence 1, Appli
19	267	90.8	110	15	US-10-038-686-1	Sequence 2, Appli
20	267	90.8	117	9	US-09-280-030-63	Sequence 63, Appli
21	267	90.8	130	9	US-09-280-030-62	Sequence 62, Appli
22	266.5	90.6	124	9	US-09-736-611-12	Sequence 12, Appli
23	266.5	90.6	124	9	US-09-740-359-12	Sequence 12, Appli
24	266.5	90.6	124	10	US-09-894-711-12	Sequence 12, Appli
25	266.5	90.6	125	9	US-09-736-611-10	Sequence 10, Appli
26	266.5	90.6	125	9	US-09-740-359-10	Sequence 10, Appli
27	266.5	90.6	125	10	US-09-894-711-10	Sequence 10, Appli
28	266.5	90.6	147	9	US-09-736-611-8	Sequence 8, Appli
29	266.5	90.6	147	9	US-09-740-359-7	Sequence 7, Appli
30	265	90.1	144	9	US-09-736-611-6	Sequence 6, Appli
31	265	90.1	144	9	US-09-740-359-5	Sequence 5, Appli
32	265	90.1	146	10	US-09-894-711-5	Sequence 5, Appli
33	261	88.9	96	10	US-09-947-563-5	Sequence 5, Appli
34	244.5	83.2	144	10	US-09-894-711-7	Sequence 7, Appli
35	214	74.5	46	9	US-09-205-658-132	Sequence 132, App
36	214	72.8	46	9	US-09-205-658-133	Sequence 133, App
37	195	68.8	46	9	US-09-205-658-135	Sequence 135, App
38	195	66.3	46	9	US-09-205-658-134	Sequence 134, App
39	194	66.0	46	9	US-09-205-658-136	Sequence 136, App
40	184	62.6	46	9	US-09-205-658-137	Sequence 137, App
41	163.5	62.4	50	9	US-09-205-658-115	Sequence 115, App
42	173	58.8	32	9	US-09-815-229-14	Sequence 14, Appli
43	173	58.8	32	10	US-09-947-563-6	Sequence 6, Appli
44	173	58.8	56	9	US-09-815-229-10	Sequence 10, Appli
45	173	58.8	166	9	US-09-925-297-805	Sequence 805, App

## ALIGNMENTS

RESULT 1  
US-10-054-873-5  
; Sequence 5, Application US/10054873  
; Publication No. US20020164712A1  
; GENERAL INFORMATION:  
; APPLICANT: Gan, Zhong Ru  
; TITLE OF INVENTION: Chimeric Protein Containing an  
; ; Intramolecular Chaperone Like Sequence  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/054,873  
; FILING DATE: 22-Jan-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/CN98/00052  
; FILING DATE: 31-MAR-1998  
; APPLICATION NUMBER: US 09/423,100  
; FILING DATE: 11-DEC-2000  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mycroft, Frank J  
; REGISTRATION NUMBER: 46,946  
; REFERENCE/DOCKET NUMBER: 020167-00013005  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 52 amino acids  
; TYPE: amino acid



```
Matches 51: Conservative 0: Mismatches 0: Indels 1: Gaps 1:
Oy 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYOLENYCN 52
Db 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYOLENYCN 124
1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYOLENYCN 51

RESULT 5
US-10-028-410-3
: Sequence 3, Application US/10028410
: Publication No. US20020160955A1
: GENERAL INFORMATION:
: APPLICANT: Dubaqui, Yves
: APPLICANT: Lowman, Henry
: TITLE OF INVENTION: PROTEIN VARIANTS
: FILE REFERENCE: P1712R1-1
: CURRENT APPLICATION NUMBER: US/10/028-410
: CURRENT FILING DATE: 2003-12-15
: PRIOR APPLICATION NUMBER: US/09/411-024
: PRIOR FILING DATE: 2003-05-05
: NUMBER OF SEQ ID NOS: 6
: SEQ ID NO: 5
: LENGTH: 51
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-028-410-3
Query Match 94.7% Score 278.5; DB 14; Length 51;
Best Local Similarity 94.3% Pred. No. 1.2e-26;
Matches 51: Conservative 0: Mismatches 0: Indels 1: Gaps 1:
Oy 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYOLENYCN 52
Db 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYOLENYCN 51

RESULT 6
US-09-894-711-18
: Sequence 18, Application US/09894711
: Patent No. US2002017144A1
: GENERAL INFORMATION:
: APPLICANT: Kjelsgaard, Thomas Borlaam
: APPLICANT: Ludvigsen, Svend
: TITLE OF INVENTION: Method for making insulin precursors and
: TITLE OF INVENTION: insulin precursor analogues having improved fermentation
: TITLE OF INVENTION: insulin precursor analogues having improved fermentation
: FILE REFERENCE: 6148.400-US
: CURRENT APPLICATION NUMBER: US/05/894-711
: CURRENT FILING DATE: 2001-06-28
: PRIOR APPLICATION NUMBER: PA 2000 00443
: PRIOR FILING DATE: 2000-03-17
: PRIOR APPLICATION NUMBER: PA 1999 01869
: PRIOR FILING DATE: 1999-12-29
: PRIOR APPLICATION NUMBER: 60/211,081
: PRIOR FILING DATE: 2000-06-13
: PRIOR APPLICATION NUMBER: 60/81,450
: PRIOR FILING DATE: 2000-02-10
: PRIOR APPLICATION NUMBER: 09/749,359
: PRIOR FILING DATE: 2000-12-19
: NUMBER OF SEQ ID NOS: 20
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO: 18
: LENGTH: 124
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Synthetic
US-09-894-711-18
Query Match 93.7% Score 275.5; DB 10; Length 124;
Best Local Similarity 94.3% Pred. No. 1.2e-26;
Matches 50: Conservative 1: Mismatches 1: Indels 1: Gaps 1:
Oy 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYOLENYCN 52
Db 86 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYOLENYCN 138

RESULT 8
US-10-066-009A-3
: Sequence 1, Application US/10066009A
: Publication No. US20020165155A1
: GENERAL INFORMATION:
: APPLICANT: Schaffer, Michelle
: APPLICANT: Uitsch, Mark
: APPLICANT: Vajdos, Felix
: TITLE OF INVENTION: CRYSTALLIZATION OF IGF-1
: FILE REFERENCE: P1869R1
```

```
Oy 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYOLENYCN 52
Db 72 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYOLENYCN 124

RESULT 7
US-09-861-687-19
: Sequence 19, Application US/09861687
: Publication No. US20020193292A2;
: GENERAL INFORMATION:
: APPLICANT: Markussen, Jan
: APPLICANT: Jonassen, Ib
: APPLICANT: Havelund, Svend
: APPLICANT: Brandt, Jakob
: APPLICANT: Karlzhals, Peter
: APPLICANT: Hansen, Heltz Per
: APPLICANT: Kaarstoft, Niels Christian
: TITLE OF INVENTION: INSULIN DERIVATIVES
: NUMBER OF SEQUENCES: 26
: REFERENCE SEQUENCES: 26
: ADDRESS: No. US20020193292A1; No. US20020193292A1; disk of No. US20020193292A1;
: STREET: 405 Lexington Avenue, 64th floor
: CITY: New York
: STATE: New York
: COUNTRY: United States of America
: ZIP: 10176 6401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/861,687
: FILING DATE: 21-May-2001
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/932,082
: FILING DATE: 16-DEC-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Lambiris, Elias J.
: REGISTRATION NUMBER: 33,728
: REFERENCE/DOCKET NUMBER: 4341,204-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-867-0123
: TELEFAX: 212-878-9655
: INFORMATION FOR SEQ ID NO: 19:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 138 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-09-861-687-19
Query Match 93.7% Score 275.5; DB 10; Length 138;
Best Local Similarity 94.3% Pred. No. 1.4e-26;
Matches 50: Conservative 1: Mismatches 1: Indels 1: Gaps 1:
Oy 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYOLENYCN 52
Db 86 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYOLENYCN 138

RESULT 8
US-10-066-009A-3
: Sequence 1, Application US/10066009A
: Publication No. US20020165155A1
: GENERAL INFORMATION:
: APPLICANT: Schaffer, Michelle
: APPLICANT: Uitsch, Mark
: APPLICANT: Vajdos, Felix
: TITLE OF INVENTION: CRYSTALLIZATION OF IGF-1
: FILE REFERENCE: P1869R1
```

; CURRENT APPLICATION NUMBER: US/10/066,009A  
; PRIOR FILING DATE: 2002-06-24  
; PRIOR APPLICATION NUMBER: US 60/287,072  
; PRIOR FILING DATE: 2001-04-27  
; PRIOR APPLICATION NUMBER: US 60/267,977  
; PRIOR FILING DATE: 2001-02-09  
; NUMBER OF SEQ ID NOS: 5  
; SEQ ID NO 3  
; LENGTH: 50  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-066-009A-3

Query Match 92.9% Score 273; DB 14; Length 50;  
Best Local Similarity 96.2% Pred. No. 9,4e-27;  
Matches 50; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRRRAEDLQVGVGLGGPGAGSLQPLALEG 52  
DB 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRRRAEDLQVGVGLGGPGAGSLQPLALEG 50

RESULT 9  
US-09-878-380-1  
; Sequence 1, Application US/09878380  
; Patent No. US20020160435A1  
; GENERAL INFORMATION:  
; APPLICANT: FUJIREBIO INC.  
; APPLICANT: KITAJIMA, Sachiko  
; APPLICANT: KURANO, Yoshihiro  
; APPLICANT: NAKATSUBO, Kaoru  
; APPLICANT: NISHIZONO, Isao  
; TITLE OF INVENTION: Immunassay For Measuring Human C-Peptide and Kit Therefor  
; FILE REFERENCE: 0760-0291P  
; CURRENT APPLICATION NUMBER: US/09/878,380  
; CURRENT FILING DATE: 2001-06-12  
; PRIOR APPLICATION NUMBER: JP 2000-174691  
; PRIOR FILING DATE: 2000-06-12  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patent In version 3.1  
; SEQ ID NO 1  
; LENGTH: 86  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-878-380-1

Query Match 90.8% Score 263; DB 10; Length 86;  
Best Local Similarity 60.5% Pred. No. 9,4e-26;  
Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRRRAEDLQVGVGLGGPGAGSLQPLALEG 60  
DB 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRRRAEDLQVGVGLGGPGAGSLQPLALEG 60

QY 31 ----RGIVEQCCTTSCISLYOLENYCN 62  
DB 61 SLQKRGIVEQCCTTSCISLYOLENYCN 60

RESULT 10  
US-09-858-935B-4  
; Sequence 4, Application US/99858945B  
; Publication No. US20030069177A1  
; GENERAL INFORMATION:  
; APPLICANT: Dubaqui, Yves  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Lowman, Henry B.  
; TITLE OF INVENTION: METHOD FOR TREATING CARTILAGE DISORDERS  
; FILE REFERENCE: P1794R1  
; CURRENT APPLICATION NUMBER: US/09/858,935B  
; CURRENT FILING DATE: 2002-07-02  
; PRIOR APPLICATION NUMBER: US 60/248,985  
; PRIOR FILING DATE: 2000-11-15

; PRIOR APPLICATION NUMBER: US 60/204,490  
; PRIOR FILING DATE: 2000-05-16  
; NUMBER OF SEQ ID NOS: 153  
; SEQ ID NO 4  
; LENGTH: 86  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-858-935B-4

Query Match 90.8% Score 267; DB 11; Length 86;  
Best Local Similarity 60.5% Pred. No. 9,4e-26;  
Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRRRAEDLQVGVGLGGPGAGSLQPLALEG 60  
DB 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRRRAEDLQVGVGLGGPGAGSLQPLALEG 60

QY 31 ----RGIVEQCCTTSCISLYOLENYCN 52  
DB 61 SLQKRGIVEQCCTTSCISLYOLENYCN 86

RESULT 11  
US-10-028-410-2  
; Sequence 2, Application US/10028410  
; Publication No. US20020160955A1  
; GENERAL INFORMATION:  
; APPLICANT: Dubaqui, Yves  
; APPLICANT: Lowman, Henry  
; TITLE OF INVENTION: PROTEIN VARIANTS  
; FILE REFERENCE: P1712P1-1  
; CURRENT APPLICATION NUMBER: US/10/028,410  
; CURRENT FILING DATE: 2001-12-19  
; PRIOR APPLICATION NUMBER: US/09/477,924  
; PRIOR FILING DATE: 2000-01-05  
; NUMBER OF SEQ ID NOS: 6  
; SEQ ID NO 2  
; LENGTH: 86  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-028-410-2

Query Match 90.8% Score 267; DB 14; Length 86;  
Best Local Similarity 60.5% Pred. No. 9,4e-26;  
Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRRRAEDLQVGVGLGGPGAGSLQPLALEG 60  
DB 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRRRAEDLQVGVGLGGPGAGSLQPLALEG 60

QY 31 ----RGIVEQCCTTSCISLYOLENYCN 52  
DB 61 SLQKRGIVEQCCTTSCISLYOLENYCN 86

RESULT 12  
US-10-054-473-4  
; Sequence 4, Application US/10054873  
; Publication No. US20020164712A1  
; GENERAL INFORMATION:  
; APPLICANT: Gan, Zhong RU  
; TITLE OF INVENTION: Chimeric Protein Containing an Intramolecular Chaperone-Like Sequence  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Ownsend and Townsend and Crow LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/054,873  
FILING DATE: 22-Jan-2002  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/CN98/00052  
FILING DATE: 31-MAR-1998  
APPLICATION NUMBER: US 09/423,100  
FILING DATE: 11-DEC-2000

ATTORNEY/AGENT INFORMATION:  
NAME: Mycroft, Frank J  
REGISTRATION NUMBER: 46,946  
REFERENCE/DOCKET NUMBER: 020167-00013005

INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 86 amino acids  
TYPE: amino acid  
STRANDEDNESS: <unknown>  
TOPOLOGY: linear

MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-10-054-873-4

Query Match 90.8%; Score 267; GR 10; Length 86;  
Best Local Similarity 60.5%; Pred. No. 1; 26;  
Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

QY 1 FVNOHLCGSHLVKALYVCGRGFFFTPKT- ----- 30

Db 1 FVNOHLCGSHLVKALYVCGRGFFFTPKT- AEDLVQGVGLGGPGAGSLOPLALEG 60

QY 31 - - - - -RGIVEQCCTSCISLYOLENYCN 52

Db 61 SLQKRGIVEQCCTSCISLYOLENYCN 86

RESULT 13  
US-09-947-563-4

Sequence 4, Application US/09947563  
Patent No. US2020156214A1  
GENERAL INFORMATION:

APPLICANT: Rubroder, Franz-Josef

TITLE OF INVENTION: Improved process for obtaining

insulin precursors having correctly bonded cystine bridges

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &

STREET: 1300 I Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005-3315

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/947,563

FILING DATE: 07-Sep-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/134,836

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Leslie McDonell

REGISTRATION NUMBER: 34,872

REFERENCE/DOCKET NUMBER: 02481.1600-00000

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 96 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Escherichia coli  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1-96  
SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-09-947-563-4

Query Match 90.8%; Score 267; GR 10; Length 96;  
Best Local Similarity 60.5%; Pred. No. 1; 26;  
Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

QY 1 FVNOHLCGSHLVKALYVCGRGFFFTPKT- ----- 30

Db 11 FVNOHLCGSHLVKALYVCGRGFFFTPKT- RREAEIDLQGVGLGGPGAGSLOPLALEG 70

QY 31 - - - - -RGIVEQCCTSCISLYOLENYCN 52

Db 71 SLQKRGIVEQCCTSCISLYOLENYCN 96

RESULT 14

US-09-205-658-125

Sequence 125, Application US/09205658

Patent No. US20010029617A1

GENERAL INFORMATION:

APPLICANT: Ruvkun, Gary

APPLICANT: Oqq, Scott

TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR

FILE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS

FILE REFERENCE: 00786/351004

CURRENT APPLICATION NUMBER: US/09/205,658

CURRENT FILING DATE: 1998-12-03

EARLIER APPLICATION NUMBER: 08/857,076

EARLIER FILING DATE: 1997-05-15

EARLIER APPLICATION NUMBER: 08/889,534

EARLIER FILING DATE: 1997-07-07

EARLIER APPLICATION NUMBER: US98/10080

NUMBER OF SEQ ID NOS: 328

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 125

LENGTH: 110

TYPE: PRT

ORGANISM: Homo sapiens

US-09-205-658-125

Query Match 90.8%; Score 267; GR 9; Length 110;

Best Local Similarity 60.5%; Pred. No. 1; 26-25;

Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

QY 1 FVNOHLCGSHLVKALYVCGRGFFFTPKT- ----- 30

Db 25 FVNOHLCGSHLVKALYVCGRGFFFTPKT- RREAEIDLQGVGLGGPGAGSLOPLALEG 84

QY 31 - - - - -RGIVEQCCTSCISLYOLENYCN 52

Db 85 SLQKRGIVEQCCTSCISLYOLENYCN 110

RESULT 15

US-09-815-229-3

Sequence 3, Application US/09815229

```

: Patent No. US20020058614A1
: GENERAL INFORMATION:
: APPLICANT: Filvaroff, Ellen H.
: APPLICANT: Okuno, Franklin W.
: TITLE OF INVENTION: USE OF INSULIN FOR THE TREATMENT OF CARTILLAGENOUS DISORDERS
: FILE REFERENCE: P1786R10S
: CURRENT APPLICATION NUMBER: US/09/815,229
: CURRENT FILING DATE: 2001-03-22
: PRIOR APPLICATION NUMBER: US 60/192,103
: PRIOR FILING DATE: 2000-01-24
: NUMBER OF SEQ ID NOS: 17
: SEQ ID NO 3
: LENGTH: 120
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-815-229-3

```

```

Query Match      90.8%   Score 267.   FR 92   Length 116
Best local similarity 60.5%   Posd No. 1 25 25
Matches 52: Conservative 0: Missed 68 0: Indels 54: Gaps 11
QY      3  FVNQLCGSHLVEALYLWGGERGFFYTKR 1
DB      25  FVNQLCGSHLVEALYLWGGERGFFYTKR 1
QY      31  ... RGIVEQCCTSIYOLENYGN 52
DB      85  SLQKRGIVEQCCTSIYOLENYGN 110

```

Search completed: September 15, 2003, 12:23:46  
Job time : 22.8065 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:56:45 : Search time 8.57348 seconds  
(without alignments)  
583.284 Million cell updates/sec

Title: US-09-423-100-5

Perfect score: 294

Sequence: 1 FVNHQJGSHLVKALYVCG.....IVFQVCTICSIVOLENYCN 52

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 283308

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	273.5	93.0	51	1 INWHP	insulin - sperm whale
2	273.5	93.0	51	1 INWHP	insulin - finback
3	273.5	93.0	51	1 INEL	insulin - elephant
4	274	92.9	96	2 PC7042	epidermal growth f
5	271.5	92.3	51	1 INHY	insulin - hamster
6	268.5	91.3	51	1 NMSSP	insulin - Egyptian
7	267.5	91.0	51	2 A59151	insulin precursor
8	267	90.8	110	1 IPHU	insulin precursor
9	267	90.8	110	2 B42179	insulin precursor
10	267	90.8	110	2 A42174	insulin precursor
11	267	90.8	110	2 JQ0178	insulin precursor
12	263.5	89.6	51	1 INWHIS	insulin - sei whal
13	263.5	89.6	51	1 INGT	insulin - goat
14	263.5	89.6	51	1 INOMA	insulin - Arabian
15	263	89.5	84	1 IPPG	insulin precursor
16	263	89.5	110	1 INRR	insulin precursor
17	262.5	89.3	110	1 INCT	insulin cat
18	262	89.3	110	1 IPDG	insulin precursor
19	261.5	88.9	51	1 INMKSU	insulin - common S
20	260	88.4	110	2 I48166	insulin precursor
21	258.5	87.9	105	1 IPBO	insulin precursor
22	256.5	87.2	51	2 JQ0162	insulin - North Am
23	252.5	85.9	77	1 INSH	insulin precursor
24	252	85.7	86	1 IPHO	insulin - Chinchil
25	251.5	85.5	51	1 INCB	insulin precursor
26	251	85.4	108	2 A39883	insulin precursor
27	250	85.0	108	1 INMS1	insulin 1 precurs
28	249	84.7	110	1 IPRT1	insulin 1 precurs
29	248.5	84.5	51	1 INGS	insulin - goose

## ALIGNMENTS

### RESULT 1

INWHP

Insulin - sperm whale

C:Species: Physeter catodon (sperm whale)

C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 16-Jul-1999

C:Accession: A93142; A90082

R: Ishihara, Y.; Saito, T.; Ito, T.; Fujino, M.

Nature 181, 1468-1469, 1958

A:Title: Structure of sperm- and sei whale insulins and their breakdown by whale pe

A:Reference number: A93142

A:Accession: A93142

A:Molecule type: protein

A:Residues: 1-30;31-51 <ISH>

R: Harris, J.I.; Sanger, F.; Naughton, M.A.

Arch. Biochem. Biophys. 65, 427-428, 1956

A:Title: Species differences in insulin.

A:Reference number: A90082

A:Accession: A90082

A:Molecule type: protein

A:Residues: 1-30;31-51 <HAR>

C:Superfamily: Insulin

C:Keywords: hormone; pancreas

F:1-30/Domain: insulin chain B #status experimental <HCH>

F:30-31-51/Product: insulin #status experimental <MAT>

F:31-51/Dxmat: insulin chain A #status experimental <ACH>

F:7-37;19-50;36-41/Bisulfide bonds: #status predicted

Query Match 93.0% Score 273.5; Dh 1; length 51;

Best local Similarity 96.2%; Pred. No. 1; 5e 24;

Matches 50; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 FVNHQJGSHLVKALYVCGERGFFYTPKTRGIVGCTICSIVOLENYCN 52

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

DB 1 FVNHQJGSHLVKALYVCGERGFFYTPKA-GIVEQCCTICSIVOLENYCN 51

### RESULT 2

INWHP

Insulin - finback whale (tentative sequence)

C:Species: Balaenoptera physalus (finback whale, common rorqual)

C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 16-Jul-1999

C:Accession: A91918

R: Hama, H.; Titani, K.; Sakaki, S.; Narita, K.

J. Biochem. 56, 285-293, 1964

A:Title: The amino acid sequence in fin-whale insulin.

A:Reference number: A91918

A:Accession: A91918

A:Molecule type: protein

A:Residues: 1-30;31-51 <HMA>

C:Superfamily: Insulin

C:Keywords: hormone; pancreas

F:1-30/Domain: insulin chain B #status experimental <BCH>



A: Molecule type: protein  
A: Residues: 31-51 <MAC>  
C: Comment: The two chains are probably produced from the same precursor.  
C: Superfamily: insulin  
F: 1-30, 31-51/Product: insulin \*status experimental <MAT>  
F: 1-30/Domain: chain B \*status experimental <HB>  
F: 31-51/Domain: chain A \*status experimental <HA>  
F: 7-37, 19-50, 36-41/Disulfide bonds: \*status predicted

Query Match 91.0%; Score 267.5; DB 2; Length 51;  
Best Local Similarity 92.3%; Pred. No. 7.2e-24;  
Matches 48; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVGCCISLSLYOLENYCN 52  
|||||  
DB 1 FVNHLCGSHLVEALYLVCGERGFFYTPKA-GIVGCCASVCSLYOLENYCN 51  
|||||

RESULT 8  
IPHU insulin precursor (validated) human  
N: Alternate names: preproinsulin  
C: Species: Homo sapiens (man)  
C: Date: 23-Oct-1981 \*sequence\_revision 23-Oct-1981 \*text\_change 08-Dec-2000  
C: Accession: A93222; A94253; A94251; A93144; A92075; A91186; 158114; A01579; S58  
R: Bell, G.I.; Pictet, R.L.; Rutter, W.J.; Luddell, B.; Tischer, E.; Goodman, H.M.  
Nature 284, 26-32, 1980  
A: Title: Sequence of the human insulin gene.  
A: Reference number: A93222; MUID: 80120725; PMID: 743748  
A: Accession: A93222  
A: Molecule type: DNA  
A: Residues: 1-110 <REL>  
A: Cross-references: GR:300265; MUID: q186429; PIDN: AAA59172.1; PID: q386828  
R: Ulrich, A.; Bull, T.J.; Gray, A.; Brostius, J.; Sures, I.  
Science 209, 612-615, 1980  
A: Title: Genetic variation in the human insulin gene.  
A: Reference number: A94253; MUID: 80236313; PMID: 6248462  
A: Accession: A94253  
A: Molecule type: DNA  
A: Residues: 1-110 <ULL>  
A: Cross-references: GR:300265; MUID: q186429; PIDN: AAA59172.1; PID: q386828  
R: Bell, G.I.; Swain, W.F.; Pictet, R.; Cordell, B.; Goodman, H.M.; Rutter, W.J.  
Nature 282, 525-527, 1979  
A: Title: Nucleotide sequence of a cDNA clone encoding human preproinsulin.  
A: Reference number: A93216; MUID: 80054779; PMID: 501234  
A: Accession: A93216  
A: Molecule type: mRNA  
A: Residues: 1-110 <REL2>  
A: Cross-references: GR:300265; MUID: q186429; PIDN: AAA59172.1; PID: q386828  
R: Sures, I.; Goedel, D.V.; Gray, A.; Ulrich, A.  
Science 208, 57-59, 1980  
A: Title: Nucleotide sequence of human preproinsulin complementary DNA.  
A: Reference number: A94251; MUID: 80147417; PMID: 6927840  
A: Accession: A94251  
A: Molecule type: mRNA  
A: Residues: 1-110 <SUR>  
A: Cross-references: GR:300265; MUID: q186429; PIDN: AAA59172.1; PID: q386828  
R: Nicol, D.S.H.; Smith, L.F.  
Nature 187, 483-485, 1960  
A: Title: Amino acid sequence of human insulin.  
A: Reference number: A93144  
A: Accession: A93144  
A: Molecule type: protein  
A: Residues: 25-54; 90-110 <NIC>  
R: Oyer, P.E.; Cho, S.; Peterson, J.D.; Steiner, D.F.  
J. Biol. Chem. 246, 1375-1386, 1971  
A: Title: Studies on human proinsulin. Isolation and amino acid sequence of the human par  
A: Reference number: A92075; MUID: 71116410; PMID: 5101771  
A: Accession: A92075  
A: Molecule type: protein  
A: Residues: 57-87 <OYE>  
R: Ko, A.; Smyth, D.G.; Markussen, J.; Sundby, F.  
Eur. J. Biochem. 20, 190-199, 1971

A: Title: Amino acid sequence of the C-peptide of human proinsulin.  
A: Reference number: A91186; MUID: 71257722; PMID: 5560404  
A: Accession: A91186  
A: Molecule type: protein  
A: Residues: 57-87 <KOA>  
R: Lucassen, A.M.; Julier, C.; Beressi, J.P.; Boitard, C.; Froguet, P.; Lathrop, M.;  
Nature Genet. 4, 305-310, 1993  
A: Title: Susceptibility to insulin dependent diabetes mellitus maps to a 4.1 kb seg  
A: Reference number: 158114; MUID: 94364428; PMID: 8358440  
A: Accession: 158114  
A: Status: preliminary; translated from GR/PMBL/DBS  
A: Molecule type: DNA  
A: Residues: 1-59, 63-110 <RES>  
A: Cross-references: GR: L15440; MUID: q307071; PIDN: AAA59179.1; PID: q307072  
R: Sieber, P.; Kamber, B.; Hartmann, A.; Jorhi, A.; Riniker, B.; Rittel, W.  
Helv. Chim. Acta 57, 2617-2621, 1974  
A: Title: Totalsynthese von Humaninsulin unter gezielter Bildung der Disulfidbindungen.  
A: Reference number: A91636; MUID: 75077277; PMID: 4434293  
A: Contents: annotation: Synthesis  
A: Note: disulfide bonded human insulin was synthesized; the synthetic hormone was 1  
A: Note: article in German with English abstract  
R: Naethan, V.K.  
Hoppe-Seyler's Z. Physiol. Chem. 354, 659-672, 1973  
A: Title: The synthesis of C-peptide of human proinsulin.  
A: Reference number: A91658; MUID: 75040007; PMID: 4803504  
A: Contents: annotation: Synthesis of residues 57-87  
R: Geiger, R.; Jaeger, G.; Koenig, W.  
Chem. Ber. 106, 2347-2352, 1973  
A: Title: Synthesis of the complete sequence of human proinsulin C-peptide and its f  
A: Reference number: A90914  
A: Contents: annotation: synthesis of residues 57-87  
R: Kaufmann, J.E.; Iminger, J.C.; Haiban, P.A.  
Biochem. J. 310, 869-874, 1995  
A: Title: Sequence requirements for proinsulin processing at the B-chain/C-peptide j  
A: Reference number: S58651; MUID: 9603185; PMID: 7575420  
A: Contents: annotation: site-directed mutagenesis study of proteolytic processing  
C: Genetics  
A: Gene: GDB: INS  
A: Cross-references: GDB: 119349; OMIM: 176730  
A: Map position: 11p15.5-11p15.5  
A: Introns: 63/1  
C: Superfamily: insulin  
C: Keywords: hormone; pancreas  
F: 1-24/Domain: signal sequence \*status predicted <SIG>  
F: 25-54/Domain: insulin chain B \*status experimental <BCH>  
F: 25-54, 90-110/Product: insulin \*status experimental <MAT>  
F: 57-87/Domain: connecting C-peptide \*status experimental <CEP>  
F: 90-110/Domain: insulin chain A \*status experimental <ACH>  
F: 31-96, 43-109, 95-100/Disulfide bonds: \*status experimental

Query Match 90.8%; Score 267; DB 1; Length 110;  
Best Local Similarity 60.5%; Pred. No. 1.6e-23;  
Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVGCCISLSLYOLENYCN 52  
|||||  
DB 25 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVGCCISLSLYOLENYCN 84  
|||||

QY 31 ----RGIVGCCISLSLYOLENYCN 52  
|||||

DB 85 SIQKRGIVGCCISLSLYOLENYCN 110  
|||||

RESULT 9  
B42179 insulin precursor - green monkey  
C: Species: Cercopithecus aethiops (green monkey, grivet)  
C: Date: 04-Mar-1993 \*sequence\_revision 18-Nov-1994 \*text\_change 16-Jul-1999  
C: Accession: B42179; A05232; S16494; S22056  
R: Seino, S.; Bell, G.I.; Li, W.H.  
Mol. Biol. Evol. 9, 193-203, 1992  
A: Title: Sequences of primate insulin genes support the hypothesis of a slower rate  
A: Reference number: A42179; MUID: 92219553; PMID: 1560757

```
A:Accession: B42179
A:Molecule type: DNA
A:Residues: 1-110 <SET>
A:Cross-references: EMBL:X61092; NID:q22808; PIDN:CAA43405.1; PID:q22809
A:Note: sequence extracted from NCBI backbone (NCBI:95185, NCBI:P:95194)
R:Peterson, J.D.; Nehrllich, S.; Over, P.E.; Steiner, D.F.
J. Biol. Chem. 247, 4866-4871, 1972
A:Title: Determination of the amino acid sequence of the monkey, sheep, and dog proinsulin
A:Reference number: A92111; MUID:72258016; PMID:4626369
A:Accession: A05232
A:Molecule type: protein
A:Residues: 57-87 <PET>
C:Genetics:
C:Introns: 63/1
C:Superfamily: insulin
C:Keywords: hormone; pancreas
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-54/Domain: insulin chain B #status predicted <BH>
F:55-87/Domain: connecting peptide #status predicted <CP>
F:90-110/Domain: insulin chain A #status predicted <AH>
F:31-96,43-109,95-100/disulfide bonds: #status predicted
Query Match 90.8%; Score 267; DB 2; Length 110;
Best Local Similarity 60.5%; Pred. No. 1 6e-23;
Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;
QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKT----- 30
DB 25 FVNHLCGSHLVEALYLVCGERGFFYTPKTRRARDPQVGVGLGGPGAGSLQPLALEG 84
QY 31 ---RGIVEQCCTSIQSLYLENYCN 52
DB 85 SLOKRGIVEQCCTSIQSLYLENYCN 110
RESULT 13
Insulin precursor - chimpanzee
C:Species: Pan troglodytes (chimpanzee)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 16-Jul-1999
A:Accession: A42179; S22058
R:Seino, S.; Bell, G.I.; Li, W.H.
Mol. Biol. Evol. 9, 193-203, 1992
A:Title: Sequences of primate insulin genes support the hypothesis of a slower rate of
A:Reference number: A42179; MUID:9221955; ID: 11560757
A:Accession: A42179
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-110 <SET>
A:Cross-references: EMBL:X61089; NID:q38251; PIDN:CAA43403.1; PID:q48252
A:Note: sequence extracted from NCBI backbone (NCBI:P:95067)
C:Genetics:
C:Introns: 63/1
C:Superfamily: insulin
Query Match 90.8%; Score 267; DB 2; Length 110;
Best Local Similarity 60.5%; Pred. No. 1 6e-23;
Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;
QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKT----- 30
DB 25 FVNHLCGSHLVEALYLVCGERGFFYTPKTRRARDPQVGVGLGGPGAGSLQPLALEG 84
QY 31 ---RGIVEQCCTSIQSLYLENYCN 52
DB 85 SLOKRGIVEQCCTSIQSLYLENYCN 110
RESULT 11
Insulin precursor - crab-eating macaque
C:Species: Macaca fascicularis (crab-eating macaque)
A:Accession: JQ0178
A:Residues: 1-110 <SET>
A:Cross-references: EMBL:X61092; NID:q22808; PIDN:CAA43405.1; PID:q22809
A:Note: sequence extracted from NCBI backbone (NCBI:95185, NCBI:P:95194)
R:Peterson, J.D.; Nehrllich, S.; Over, P.E.; Steiner, D.F.
J. Biol. Chem. 247, 4866-4871, 1972
A:Title: Determination of the amino acid sequence of the monkey, sheep, and dog proinsulin
A:Reference number: A92111; MUID:72258016; PMID:4626369
A:Accession: A05232
A:Molecule type: protein
A:Residues: 57-87 <PET>
C:Genetics:
C:Introns: 63/1
C:Superfamily: insulin
C:Keywords: hormone; pancreas
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-54/Domain: insulin chain B #status predicted <BH>
F:55-87/Domain: connecting peptide #status predicted <CP>
F:90-110/Domain: insulin chain A #status predicted <AH>
F:31-96,43-109,95-100/disulfide bonds: #status predicted
Query Match 90.8%; Score 267; DB 2; Length 110;
Best Local Similarity 60.5%; Pred. No. 1 6e-23;
Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;
QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKT----- 30
DB 25 FVNHLCGSHLVEALYLVCGERGFFYTPKTRRARDPQVGVGLGGPGAGSLQPLALEG 84
QY 31 ---RGIVEQCCTSIQSLYLENYCN 52
DB 85 SLOKRGIVEQCCTSIQSLYLENYCN 110
RESULT 13
Insulin - goat
C:Species: Capra aegagrus hircus (domestic goat)
C:Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 16-Jul-1999
A:Accession: A01586
R:Smith, I.F.
Am. J. Med. 40, 662-666, 1966
A:Title: Species variation in the amino acid sequence of insulin.
A:Reference number: A90029; MUID:66160119; PMID:5949593
A:Accession: A01586
A:Molecule type: protein
A:Residues: 1-30:31-51 <SMI>
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:54:00 ; Search time 4.84588 seconds  
(without alignments)  
504.633 Million cell updates/sec

Title: US-09-423-100-5

Perfect score: 294

Sequence: 1 FVNQHICGSHLVEALYLVCGR.....IVEGCTSTCSLYOLENYCN 52

Scoring table: BLOSUM62

Gapof 10.0 , Gapext 0.5

Searched: 12786.4 seqs, 479,263,95 residues

Total number of hits satisfying chosen parameters: 12786.4

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	273.5	93.0	51	INS_RALPH	P01312 balaenopter
2	273.5	93.0	51	INS_ELMA	P01316 elephas max
3	268.6	91.3	51	INS_ACOCA	P03324 acromys cabi
4	267	90.8	110	INS_CERAE	P04007 ceropithec
5	267	90.8	110	INS_HUMAN	P01308 homo sapien
6	267	90.8	110	INS_MACEA	P30406 macaca fasc
7	267	90.8	110	INS_PANTR	P30410 pan troglod
8	263.5	89.6	51	INS_RALHO	P01314 balaenopter
9	263.5	89.6	51	INS_CAMDR	P01320 camelus dro
10	264.5	89.6	51	INS_CAPHI	P01319 capra hircu
11	263	89.5	108	INS_PIG	P01315 sus scrofa
12	263	89.5	110	INS_RHIT	P01311 oryctolagus
13	263	89.5	110	INS_SPERH	G91X13 spermophilu
14	262.5	89.3	51	INS_FELCA	P06306 felis silve
15	262	89.3	110	INS_CANFA	P01321 canis fami
16	260	88.4	110	INS_CRILO	P01313 cricetus
17	258.5	87.9	105	INS_BOVIN	P01317 bos taurus
18	257	87.4	110	INS_PCAOR	G62587 psammomys c
19	256.5	87.2	51	INS_DUMA	P01309 didelphis m
20	254.5	86.6	105	INS_SHEEP	P01318 ovis aries
21	252	85.7	86	INS_HORSE	P01310 equus cabal
22	251.5	85.5	51	INS_CHIRR	P01327 chinchilla
23	251	85.4	108	INS_AOTTR	P10604 aotus trivi
24	250	85.0	108	INS_MOUSE	P01325 mus musculu
25	249	84.7	110	INS1_RAT	P01322 rattus norv
26	248.5	84.5	51	INS1_ANAN	P07454 anser anser
27	248	84.4	110	INS2_MOUSE	P01326 mus musculu
28	248	84.4	110	INS2_RAT	P01323 rattus norv
29	246	83.7	52	INS1_ACIU	P81423 acipenser g
30	244.5	83.2	51	INS_HYSCR	P01328 hystrix cri
31	244.5	83.2	51	INS1_TRASC	P31887 trachemys s
32	238.5	81.1	103	INS1_SELRF	P51463 selasphorus
33	235.5	80.1	51	INS_ORNAN	O91qy7 ornithorhyn

RESULT 1

INS_RALPH	107	INS_CHICK	34	234.5	79.8
AC P01312	81	INS_ANAPL	35	233.5	79.4
D: 21-JUL-1986 (Rel. 01, created)	51	INS_ALLMI	36	231.5	78.7
D: 21-JUL-1986 (Rel. 01, last sequence update)	52	INS_LEPSP	37	231	78.6
D: 01-OCT-1996 (Rel. 34, last annotation update)	51	INS_ZAODH	38	228.5	77.7
DE Insulin.	51	INS_ZAODH	39	227.5	77.4
GN INS.	106	INS1_XENLA	40	226.5	77.0
OS Balaenoptera physalus (Finback whale) (Common rorqual), and	106	INS1_XENLA	41	226.5	77.0
OS Physeter catodon (Sperm whale) (Physeter macrocephalus).	106	INS1_XENLA	42	222	75.5
OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:	57	INS_PETMA	43	216	73.5
OC Mammalia: Eutheria: Cetartiodactyla: Cetacea: Mysticeti:	51	INS_PLAFA	44	214	72.8
OC Balaenopteridae: Balaenoptera.	52	INS_AMICA	45	211	71.8
OX NCBI_TaxID=9770, 9755;	50	INS_ONCGO			
RP [1]					
RP PARTIAL SEQUENCE.					
RC SPECIES-B.physalus;					
RA Hama H., Titani K., Sakaki S., Narita K.;					
RT "The amino acid sequence in fin-whale insulin.";					
KL J. Biochem. 56:285-293(1964).					
RN [2]					
RC SPECIES-P.catodon;					
RP SPECIES-P.catodon;					
RA Ishihara Y., Saito T., Ito Y., Fujino M.;					
RT "Structure of sperm- and sei-whale insulins and their breakdown by					
RL whale pepsin.";					
RN Nature 181:1468-1469(1958).					
RN [3]					
RC SPECIES-P.catodon;					
RP SPECIES-P.catodon;					
RA Harris J. I., Sanger F., Naughton M. A.;					
RT "Species differences in insulin.";					
KL Arch. Biochem. Biophys. 65:427-448(1956).					
CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT					
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND					
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE					
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.					
CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO					
CC DISULFIDE BONDS.					
CC -!- SUBCELLULAR LOCATION: Secreted.					
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.					
DR PIR: A91918; IWHNF.					
DR PIR: A91142; IWHNF.					
DR HSSP: P01317; IAPH.					
DR InterPro: IPR024825; Ins/IGF/relax.					
DR SMART: SM00098; IIGF.1.					
DR PROSITE: PS00262; INSULIN; 1					
DR Insulin family; Hormone; Glu					
DR CHAIN 1 30					
DR CHAIN 30 31					
DR CHAIN 31 51					
DR DISULFID 7 37					
DR DISULFID 19 50					
DR DISULFID 36 41					
DR SEQUENCE 51 AA; 5766 MW; 9007B514691A7CD; CRC64;					

P01332 gallus gall  
P01333 anas platyr  
P12703 alligator m  
P09476 igipossteus  
P12708 zaocys dhum  
P01334 crotalus at  
P12706 xenopus lae  
P12707 xenopus lae  
P14806 petromyzon  
P09477 platichthys  
P29335 amia calva  
P21187 oncorhynchus

```

OX NCBI_TaxID=10068;
RN [1]
RN COMPOSITION.
RX MEDLINE=72189454; PubMed=5028210;
RA Buzenli H.F., Humbej R.E.;
RT "Isolation and partial structural analysis of insulin from mouse (Mus musculus) and spiny mouse (Acomys cahirinus).";
RL Hoppe-Seyler's Z. Physiol. Chem. 353:444-450(1972).
CC -I- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -I- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO DISULFIDE BONDS.
CC -I- SURCELLULAR LOCATION: Secreted.
CC -I- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR PER. A01591; INNSPP
DR HSD. P01408; IYIM
DR INTERPRO: IPR004876; IPR013727; IPR013728
DR SMART: SM00074; IIC001
DR PROSITE: PS00262; INSULIN_1
KW Insulin family; Hormone; Glucose metabolism; SM.
FT CHAIN 1 59 INSULIN B CHAIN.
FT NON_CONS 30 51
FT FT CHAIN 33 53 INTERCHAIN (BY SIMILARITY).
FT DISULFID 7 7 INTERCHAIN (BY SIMILARITY).
FT DISULFID 19 51 BY SIMILARITY.
FT DISULFID 36 41
SQ SEQUENCE 51 AA: 5748 MW: 49240RB625647D3D PR664;
Query Match: 91.3%; Score 268.5; DB 1; Length 51;
Best Local Similarity 92.1%; Pred. No. 26-26;
Matches 48; Conservative 3; Mismatches 0; Indels 1; Gaps
OY 1 FVNHLCGSHLVIALV:VGGEGFFYTPKRCIVGQCTCSLYQLNVCN 52
DB 1 FVNHLCGSHLVIALV:VGGEGFFYTPKRCIVGQCTCSLYQLNVCN 51
RESULT 4
INS_CERAF
ID INS_CERAF STANDARD; PRI: 110 AA.
AC P30407; P01309;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Insulin precursor.
GS INS.
QS Cercopithecus aethiops (Green monkey) (Olivet).
OC Eukaryota; Metazoa; Chordata; Cladonia; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxID=9514;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=92215953; PubMed=1560757;
RA Selino S., Hall G.J., Li W.;
RT "Sequences of primate insulin genes support the hypothesis of a slower rate of molecular evolution in humans and apes than in monkeys.";
RL Mol. Biol. Evol. 9:193-203(1992).
RN [2]
RN SEQUENCE OF 57-87.
RX MEDLINE=72258016; PubMed=4626369;
RA Peterson J.D., Nehrlich S., Oyer P.E., Steiner D.F.;
RT "Determination of the amino acid sequence of the monkey, sheep, and dog proinsulin C-peptides by a semi-micro Edman degradation procedure.";
RL J. Biol. Chem. 247:4866-4871(1972).
CC -I- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.

```

```

CC CC      1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC CC      DISULFIDE BONDS.
CC CC      1- SURCELLULAR LOCATION: Secreted.
CC CC      1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC CC      -----
CC CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC      the European Bioinformatics Institute. There are no restrictions on its
CC CC      use by non-profit institutions as long as its content is in no way
CC CC      modified and this statement is not removed. Usage by and for commercial
CC CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC      or send an email to license@isb-sib.ch).
CC CC      -----
CC CC      EMBL: X61092; CAA43405.1; .
CC CC      DR PIR: R42179; R42179.
CC CC      DR HSSP: P01308; 1A10.
CC CC      DR InterPro: IPR004825; Ins/IGF/relax.
CC CC      DR Pfam: PF00494; Ins100.1.
CC CC      DR SMART: SM00498; IGF1.
CC CC      DR PROSITE: PS00262; INSULIN; 1.
CC CC      KW Insulin family; Hormone; Glucose metabolism; Signal.
CC CC      FT SIGNAL: 1 24
CC CC      FT CHAIN: 25 54 INSULIN B CHAIN.
CC CC      FT PROPEP: 57 87 C PEPTIDE.
CC CC      FT CHAIN: 90 110 INSULIN A CHAIN.
CC CC      FT DISULFID: 31 96 INTERCHAIN.
CC CC      FT DISULFID: 43 109 INTERCHAIN.
CC CC      FT DISULFID: 95 100
CC CC      SQ SEQUENCE 110 AA: 95A1F54BF7B247F9 CRC64;
CC CC
CC CC      Query Match 90.88; Score 267; DP 1; Length 110;
CC CC      Best Local Similarity 60.58; Pred. No. 6 30-26;
CC CC      Matches 52; Conservative 0; Mismatches 6; Indels 14; Gaps 1;
CC CC
CC CC      QY 1 FYNQHLGSHLVEAIVLCGERGFFYTPKTI 30
CC CC      11111111111111111111111111111111
CC CC      DB 25 FYNQHLGSHLVEAIVLCGERGFFYTPKTI KEADHVGQVGLGGPGAGSLQPLALEG 84
CC CC
CC CC      QY 31 -RGIVEQCCTSCISLYOLENYCN 52
CC CC      11111111111111111111111111111111
CC CC      DB 85 SLOKRGIVEQCCTSCISLYOLENYCN 110
CC CC
CC CC      RESULT 5
CC CC      INS_HUMAN
CC CC      ID INS_HUMAN STANDARD: PR3: 11; AA.
CC CC      AC P01308;
CC CC      DT 21-JUL-1986 (Ref. 01, Created)
CC CC      DT 21-JUL-1986 (Ref. 01, Last sequence update)
CC CC      DT 15-SEP-2003 (Ref. 42, Last annotation update)
CC CC      DE Insulin precursor.
CC CC      GN INS.
CC CC      OS Homo sapiens (human).
CC CC      OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC CC      OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CC CC      OX NCBI_TaxID:9606;
CC CC      RN [1]
CC CC      RP SEQUENCE FROM N.A.
CC CC      RX MEDLINE-80120725; PubMed-6243748;
CC CC      RA Bell G.I., Pictet R.L., Rutter W.J., Cordell B., Tischer E.,
CC CC      RA Goodman H.M.;
CC CC      RT "Sequence of the human insulin gene."
CC CC      RL Nature 284:26-32(1980).
CC CC      RN [2]
CC CC      RP SEQUENCE FROM N.A.
CC CC      RX MEDLINE-80236313; PubMed-6248962;
CC CC      RA Ullrich A., Dull T.J., Gray A., Brosius J., Sures I.;
CC CC      RT "Genetic variation in the human insulin gene."
CC CC      RL Science 209:612-615(1980).
CC CC      RN [3]
CC CC      RP SEQUENCE FROM N.A.
CC CC      RX MEDLINE-80054779; PubMed-503234;
CC CC      RA Bell G.I., Swain W.F., Pictet R.L., Cordell B., Goodman H.M.,

```

```

RA Rutter W.J.;
RA "Nucleotide sequence of a cDNA clone encoding human preproinsulin."
RA Nature 282:525-527(1979).
RA [4]
RA SEQUENCE FROM N.A.
RA MEDLINE-80147417; PubMed-6927840;
RA Sures I., Goeddel D.V., Gray A., Ullrich A.;
RA "Nucleotide sequence of human preproinsulin complementary DNA."
RA Science 208:57-59(1980).
RA [5]
RA SEQUENCE FROM N.A.
RA MEDLINE-93364424; PubMed-8458440;
RA Lucassen A.M., Reil J.I., Julier C., Bathrop M.;
RA "Susceptibility to insulin dependent diabetes mellitus maps to a 4.1
RA kb segment of DNA spanning the insulin gene and associated VNTR."
RA Nat. Genet. 4:305-310(1993).
RA [6]
RA SEQUENCE FROM N.A.
RA TISSUE-Pancreas;
RA MEDLINE-22348252; PubMed-12477342;
RA Sittausberg K.L., Fritzsche E.A., Grosse H.B., Larue J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.B.,
RA Altschul S.F., Zuberq B., Buetow K.E., Schaffer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Fairer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Ronald M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loughran N.A., Peters G., Abramson R.D., Mullaly S.J.,
RA Hosak S.A., McEwan P.J., McKernan K., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,
RA Villalon D.K., Mizny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.L., Skalska U., Small D.E.,
RA Schurch A., Schein J.E., Jones S.J.M., Maitra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length
RA human and mouse cDNA sequences."
RA Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RA [7]
RA SEQUENCE OF 1-59 FROM N.A.
RA TISSUE-Blood;
RA Fajardy I.I., Weill J.J., Stuckens C.C., Danze P.M.P.;
RA "Description of a novel RFLP diallelic polymorphism (-127 BsqI C/G)
RA within the 5' region of insulin gene."
RA Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RA [8]
RA SEQUENCE OF 25-54 AND 90-110.
RA Nicol D.S.H.W., Smith L.F.;
RA "Amino-acid sequence of human insulin."
RA Nature 187:483-485(1960).
RA [9]
RA SEQUENCE OF 57-87.
RA MEDLINE-71116410; PubMed-5101771;
RA Oyer P.E., Cho S., Peterson J.D., Steiner D.F.;
RA "Studies on human proinsulin. Isolation and amino acid sequence of
RA the human pancreatic C-peptide."
RA J. Biol. Chem. 246:1375-1386(1971).
RA [10]
RA SEQUENCE OF 57-87.
RA MEDLINE-71257722; PubMed-5560404;
RA KO A., Smyth D.G., Markussen J., Sundby F.;
RA "The amino acid sequence of the C-peptide of human proinsulin."
RA Eur. J. Biochem. 26:190-199(1971).
RA [11]
RA SYNTHESIS.
RA MEDLINE-75077277; PubMed-4443293;
RA Sieber P., Kamber B., Hartmann A., Joehl A., Riniker B., Rittel W.;
RA "Total synthesis of human insulin under directed formation of the
RA disulfide bonds."
RA Helv. Chim. Acta 57:2617-2621(1974).
RA [12]

```



the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: J00336; AAA36849.1; .  
 PIR: J00178; J00178.  
 HSSP: P01308; IAI0.  
 InterPro: IPR004825; Ins/IGF/relax.  
 Pfam: PF00049; Insulin; 1.  
 PRINTS: PR00277; INSULIN.  
 SMART: SM00078; IIGF; 1.  
 PROSITE: PS00262; INSULIN; 1.  
 Insulin family; Hormone; Glucose metabolism; Signal: 3D-structure.  
 FT SIGNAL 1 24  
 FT CHAIN 25 54 INSULIN B CHAIN.  
 FT PROPEP 57 87 C PEPTIDE.  
 FT CHAIN 90 110 INSULIN A CHAIN.  
 FT DISULFID 31 96 INTERCHAIN.  
 FT DISULFID 43 109 INTERCHAIN.  
 FT DISULFID 95 100  
 SQ SEQUENCE 110 AA; 11991 MW; 836E33A80A420F9 CRC64;

Query Match 90.8%; Score 267; DB 1; Length 110;  
 Best Local Similarity 60.5%; Pred. No. 6,3e-26;  
 Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;  
 QY 1 FVNHLCGSHLVEALYLVCGERGFFYPTK----- 30  
 DB 25 FVNHLCGSHLVEALYLVCGERGFFYPTKTRKAEADPQGVQVELGGSPGAGSLQPLALEG 84  
 QY 31 ---RGIVQGCCTSCSYLYENYCN 52  
 DB 85 SLQKRGIVQGCCTSCSYLYENYCN 110

RESULT 7  
 INS\_PANTR  
 ID INS\_PANTR STANDARD; PRT: 110 AA.  
 AC P30410;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Insulin precursor.  
 GN INS.  
 OS Pan troglodytes (Chimpanzee).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
 OX NCBI\_TaxID=9598;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92214453; PubMed=1560757;  
 RA Seino S., Bell G.I., Li W.;  
 RT "Sequences of primate insulin genes support the hypothesis of a  
 RT slower rate of molecular evolution in humans and apes than in  
 RT monkeys".  
 RL Mol Biol. Evol. 9:193-203(1992).  
 CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
 CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
 CC DISULFIDE BONDS.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: X61089; CAA43403.1; .  
 PIR: A42179; A42179.  
 PDB: 1EFE; 29-MAR-00.  
 InterPro: IPR004825; Ins/IGF/relax.  
 Pfam: PF00049; Insulin; 1.  
 PRINTS: PR00277; INSULIN.  
 SMART: SM00078; IIGF; 1.  
 PROSITE: PS00262; INSULIN; 1.  
 Insulin family; Hormone; Glucose metabolism; Signal: 3D-structure.  
 FT SIGNAL 1 24  
 FT CHAIN 25 54 INSULIN B CHAIN.  
 FT PROPEP 57 87 C PEPTIDE.  
 FT CHAIN 90 110 INSULIN A CHAIN.  
 FT DISULFID 31 96 INTERCHAIN.  
 FT DISULFID 43 109 INTERCHAIN.  
 FT DISULFID 95 100  
 SQ SEQUENCE 110 AA; 12025 MW; 41EB80F7963DEFF5 CRC64;

Query Match 90.8%; Score 267; DB 1; Length 110;  
 Best Local Similarity 60.5%; Pred. No. 6,3e-26;  
 Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;  
 QY 1 FVNHLCGSHLVEALYLVCGERGFFYPTK----- 30  
 DB 25 FVNHLCGSHLVEALYLVCGERGFFYPTKTRKAEADPQGVQVELGGSPGAGSLQPLALEG 84  
 QY 31 ---RGIVQGCCTSCSYLYENYCN 52  
 DB 85 SLQKRGIVQGCCTSCSYLYENYCN 110

RESULT 8  
 INS\_BALBO  
 ID INS\_BALBO STANDARD; PRT: 51 AA.  
 AC P01314;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Insulin.  
 GN INS.  
 OS Balanoptera borealis (Sei whale).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;  
 OC Balanopteridae; Balanoptera.  
 OX NCBI\_TaxID=9768;  
 RN [1]  
 RP SEQUENCE.  
 RA Ishihara Y., Saito T., Ito Y., Fujino M.;  
 RT "Structure of sperm- and sei-whale insulins and their breakdown by  
 RT whale pepsin".  
 RL Nature 181:1468-1469(1958).  
 CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
 CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
 CC DISULFIDE BONDS.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 CC PIR: A01582; INWHIS.  
 HSSP: P01317; IAPH.  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR SMART: SM00078; IIGF; 1.  
 DR PROSITE: PS00262; INSULIN; 1.  
 Insulin family; Hormone; Glucose metabolism.  
 FT CHAIN 1 30 INSULIN B CHAIN.  
 FT NON\_CONS 30 31  
 FT CHAIN 31 51 INSULIN A CHAIN.  
 FT DISULFID 7 37 INTERCHAIN.  
 FT DISULFID 19 50  
 FT DISULFID 36 41



RP REVISION TO 59.  
RA Chance R.E.;  
RL Submitted (JUL-1970) to the PIR data bank.  
[5]  
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
RA Blundell T.L., Dodson G.G., Hodgkin D., Mercola D.;  
RT Insulin. The structure in the crystal and its reflection in chemistry and biology.;  
RT Adv. Protein Chem. 26:279-402(1972).  
RL [6]  
RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).  
RA Isaacs N.W., Agarwal R.C.;  
RT Experience with fast Fourier least squares in the refinement of the crystal structure of rhombohedral 2-zinc insulin at 1.5-A resolution.;  
RL Acta Crystallogr. A 34:782-791(1978).  
RN [7]  
RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).  
RX MEDLINE: 8909418; PubMed: 2465486;  
RA Baker E.N., Blundell T.L., Cutfield J.P., Cutfield J.P., Dodson G.G., Dodson G.G., Crowfoot Hodgkin D.M., Hubbard R.E., Isaacs N.W., Reynolds C.D., Sakabe K., Sakabe N., Vijayan N.M.;  
RT The structure of 22n pig insulin crystals at 1.5 A resolution.;  
RL Philos. Trans. R. Soc. Lond., B, Biol. Sci. 319:369-456(1988).  
RN [8]  
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
RX MEDLINE: 92126280; PubMed: 1772633;  
RA Balschmidt P., Hansen F.B., Dodson E., Dodson G., Korbner F.;  
RT Structure of porcine insulin co-crystallized with glupeine Z.;  
RL Acta Crystallogr. B 47:975-986(1991).  
RN [9]  
RP X-RAY CRYSTALLOGRAPHY.  
RX MEDLINE: 91222450; PubMed: 2025410;  
RA Badger J., Harris M.R., Reynolds G.D., Evans A.C., Dodson E.;  
RT Structure of the pig insulin dimer in the cubic crystal.;  
RL Acta Crystallogr. B 47:127-136(1991).  
RN [10]  
RP X-RAY CRYSTALLOGRAPHY (1.65 ANGSTROMS).  
RA Gao J.-S., Wan Z.-L., Chang W.-R., Jiang D.-C.;  
RT Structure of monomeric porcine DesB1-B2 desptapeptide (R26-B30) insulin at 1.65 A resolution.;  
RL Acta Crystallogr. D 51:507-512(1997).  
CC [11] FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION.;  
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
CC [12] SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO DISULFIDE BONDS.  
CC [13] SUBCELLULAR LOCATION: Secreted.  
CC [14] SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
CC [15] DATABASE: NAME-Protein Spotlight;  
CC NOTE-Issue 9 of April 2003;  
CC WWW: <http://www.expasy.org/spotlight/articles/sptlr009.html>.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AF064555; AAC77920.1; ALT\_INIT.  
CC FMBL: AY044828; AAL69550.1;  
CC PDB: 3INS; 09-JAN-89.  
CC PDB: 4INS; 31-JUL-94.  
CC PDB: 6INS; 31-JAN-94.  
CC PDB: 7INS; 31-JAN-94.  
CC PDB: 9INS; 15-OCT-91.  
CC PDB: 11ZB; 15-OCT-91.  
CC PDB: 112B; 15-OCT-91.  
CC PDB: 2TC1; 29-JAN-96.

DR PDB: 1MPJ; 29-JAN-96.  
DR PDB: 3MTH; 29-JAN-96.  
DR PDB: 10EI; 16-JUN-97.  
DR PDB: 1SDB; 01-APR-98.  
DR PDB: 1WAV; 28-FEB-97.  
DR PDB: 1ZBI; 16-FEB-99.  
DR PDB: 1ZNI; 28-JAN-98.  
DR PDB: 1ZNI; 28-JAN-98.  
DR InterPro: IPR004825; Ins/IGF/relax.  
DR Pfam: PF00049; Insulin; 1.  
DR SMART: SM00078; IIGF; 1.  
DR PROSITE: PS00262; INSULIN; 1.  
KW Insulin family; Hormone; Glucose metabolism; Signal; 3D structure.  
FT SIGNAL 1 24  
FT CHAIN 25 54 INSULIN A CHAIN.  
FT PROPEP 57 85 C PEPTIDE.  
FT CHAIN 86 108 INSULIN A CHAIN.  
FT CHAIN 86 108 INTERCHAIN.  
FT DISULFID 93 98 INTERCHAIN.  
FT DISULFID 93 98 INTERCHAIN.  
FT HELIX 26 45  
FT STRAND 48 48  
FT HELIX 89 94  
FT HELIX 100 106  
FT STRAND 107 107  
FT SEQUENCE 108 AA; 11671 MW; CB4491M429H5REBE CRC64;  
SQ  
Query Match 89.5% Score 263; DR 1; Length 108;  
Best Local Similarity 60.7% Prod No 190-25;  
Matches 51; Conservative 0; Mismatches 1; Indels 32; Gaps 1;  
QY 1 FVNQHDSHLSVEALYLVCGERGFFYTPKT-- ..... 40  
DB 25 FVNQHDSHLSVEALYLVCGERGFFYTPKTARAEAFNPOAGAVELGGIGLQALALEGPP 84  
QY 31 ---RGIVEGCTISCSLYQLENYCN 52  
DB 85 QRGIVEGCTISCSLYQLENYCN 106  
RESULT: 12  
INS\_RABIT  
ID INS\_RABIT STANDARD; PRT; 110 AA.  
AC P01311;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DE 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Insulin precursor.  
GN INS.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus  
OX NCBI\_TaxID:9986;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-New Zealand white; TISSUE-Pancreas;  
RX MEDLINE:94179230; PubMed:8132571;  
RA Devaskar S.U., Giddings S.J., Rajakumar P.A., Carnaghi L.R., Menon R.K., Zahm D.S.;  
RT Insulin gene expression and insulin synthesis in mammalian neuronal cells.;  
RT J. Biol. Chem. 269:8445-8454(1994).  
RN [2]  
RP SEQUENCE OF 25-54 AND 90-110.  
RX MEDLINE:66160119; PubMed:5946593;  
RA Smith L.F.;  
RT Species variation in the amino acid sequence of insulin.;  
RL Am. J. Med. 40:662-666(1966).  
RN [3]  
RP SEQUENCE OF 56-110 FROM N.A.  
RA Giddings S.J., Carnaghi L.R., Devaskar S.U.;  
RL Submitted (APR-1991) to the EMBL/GenBank/DBJ databases  
CC [4] FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:54:30 : Search time 20.3154 Seconds  
(without alignments)  
640,520 Million cell updates/sec

Title: US-09-423-100-5

Perfect score: 294

Sequence: 1 FVNQHLCGSHLVEALYLVCGR.....IVEQCCTSIICSLYLENYCN 52

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 200525 seqs, 26405604 residues

Total number of hits satisfying chosen parameters: 4602

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP\_archaea.\*  
2: SP\_bacteria.\*  
3: SP\_fungi.\*  
4: SP\_human.\*  
5: SP\_invertebrate.\*  
6: SP\_mammal.\*  
7: SP\_mhc.\*  
8: SP\_organelle.\*  
9: SP\_plage.\*  
10: SP\_plant.\*  
11: SP\_rodent.\*  
12: SP\_virus.\*  
13: SP\_vertebrate.\*  
14: SP\_unclassified.\*  
15: SP\_virus.\*  
16: SP\_bacterioid.\*  
17: SP\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	267	90.8	110	6	Q8HXV2
2	263	89.5	110	11	Q91X14
3	251	85.4	110	6	Q8WNR6
4	219.5	74.7	106	14	Q41R07
5	201.5	68.5	110	13	Q98TA8
6	195.5	66.5	108	13	Q9D0F5
7	195.5	66.5	108	13	Q90ZM4
8	195	66.3	111	13	Q98160
9	193.5	65.8	110	13	Q90ZY1
10	191.5	65.1	111	13	Q98TA7
11	187.5	63.8	87	13	Q98TA9
12	186.5	63.4	108	13	Q98T81
13	185.5	63.1	91	13	Q98T81
14	146	49.7	65	6	Q8H281
15	146	49.7	65	6	Q8H280
16	144	49.0	207	13	Q90XD0

17	144	49.0	215	13	073721
18	143.5	48.8	132	13	Q8AV14
19	141.5	48.1	159	13	Q93607
20	141.5	48.1	182	13	073720
21	141.5	48.1	182	13	042289
22	141.5	48.1	182	13	P79824
23	141.5	48.1	185	13	057436
24	141.5	48.1	186	13	093527
25	141	48.0	185	13	Q9V157
26	140.5	47.8	176	13	Q91161
27	140.5	47.8	117	13	Q91476
28	140.5	47.8	145	13	Q95475
29	140.5	47.8	149	13	Q91231
30	140.5	47.8	155	13	Q91162
31	140.5	47.8	161	13	Q91230
32	140.5	47.8	184	13	042336
33	140.5	47.8	188	13	P81268
34	140.5	47.8	188	13	Q91465
35	140	47.6	210	13	Q91443
36	139	47.3	212	13	Q81164
37	138.5	47.1	216	13	Q42429
38	136.5	47.1	174	6	Q8M115
39	136.5	46.6	174	13	Q93389
40	137.5	46.6	62	13	Q91AA0
41	137.5	46.8	117	13	Q91914
42	137.5	46.8	149	6	Q9YX4
43	137.5	46.8	163	13	Q90VV9
44	137.5	46.8	161	13	Q9PWK2
45	137.5	46.8	161	13	Q98SR6

#### ALIGNMENTS

##### RESULT 1

Q8HXV2 PRELIMINARY: PRT: 110 AA.  
AC Q8HXV2: 01-MAR-2003 (TRENBIrel. 23, Created)  
DT 01-MAR-2003 (TRENBIrel. 23, Last sequence update)  
DT 01-MAR-2003 (TRENBIrel. 23, Last annotation update)  
DE Insulin precursor.  
GN INS.  
OS Pongo pygmaeus (Grauquana).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthalia; Primates; Catarrhini; Hominiidae; Hominidae; Pongo.  
OX NCBI\_TaxID=9600;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Stead J.D.H., Jeffreys A.J.;  
RT "Haplotype diversity at the insulin region."  
RL Submitted (Jul-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AY137503; AAN06917; J.  
SQ SEQUENCE 110 AA: 12038 MW: 2202832B94F520F8 CRC64;

Query Match 90.8% Score 267: DB 6: Length 110:  
Best Local Similarity 60.5% Pred. No. 16-28:  
Matches 52: Conservative 0: Mismatches 0: Indels 34: Gaps 1:

QY 1 FVNQHLCGSHLVEALYLVCGRGFFVTPKPT----- 30  
DB 25 FVNQHLCGSHLVEALYLVCGRGFFVTPKPTKRAFLQVQVLELGGIPNAGSLQPLALEG 84  
QY 31 -----RGIVEQCCTSIICSLYLENYCN 52  
DB 85 SLQKRGIVEQCCTSIICSLYLENYCN 110

##### RESULT 2

Q91X13 PRELIMINARY: PRT: 110 AA.  
ID Q91X13:  
AC Q91X13:  
DT 01-DEC-2001 (TRENBIrel. 19, Created)

DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
 DE Insulin.  
 OS *Spermophilus tridecemlineatus* (Thirteen-lined ground squirrel).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;  
 CC Spermophilus.  
 OX NCRI\_TaxID:43179;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreas;  
 RA Tiedora M.W., Buck M.J., Gubanoyi G., Squire T.L., Andrews M.T.;  
 RT "Regulation of PK4 expression in a hibernating mammal."  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.  
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 DR EMBL: AY048694; AAK72568.1;  
 DR HSSP: P01308; 1NF;  
 DR InterPro: IPR04825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin.  
 DR PRINTS: PK00277; INSULIN.  
 DR SMART: SM00078; IIGF; 1.  
 DR PROSITE: PS00262; INSULIN; 1.  
 SQ SEQUENCE 110 AA: 12004 MW: 4511768.662ZBEE5 CRC64:

Query Match 89.5%; Score 263; DB 11; Length 110;  
 Best Local Similarity 59.3%; Pred. No. 4.6e-28;  
 Matches 59; Conservative 1; Mismatches 0; Indels 14; Gaps 1;  
 QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTR----- 30  
 DB 25 FVNHLCGSHLVEALYLVCGERGFFYTPKSRREVFVGGQVGLGPGAGLQPLALEM 84  
 QY 31 ----RGIVEQCCTSIQSLEYQENYCN 52  
 DB 85 ALQKRGIVEQCCTSIQSLEYQENYCN 110

RESULT 3  
 Q8WNW6 PRELIMINARY: PRT: 110 AA.  
 AC Q8WNW6;  
 DT 01-MAR-2002 (TRENBLrel. 20, Created)  
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 21, Last annotation update)  
 DE Preproinsulin.  
 OS *Felis silvestris catus* (Cat).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.  
 OX NCRI\_TaxID:9685;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreas;  
 RA Okamoto S., Morimatsu M.;  
 RT "cat insulin."  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.  
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 DR EMBL: AB043535; BAB84110.1;  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1.  
 DR PRINTS: PK00277; INSULIN.  
 DR SMART: SM00078; IIGF; 1.  
 DR PROSITE: PS00262; INSULIN; 1.  
 SQ SEQUENCE 110 AA: 12069 MW: 95FB6E170C7B8CA4 CRC64:

Query Match 85.4%; Score 251; DB 6; Length 110;  
 Best Local Similarity 55.8%; Pred. No. 1.6e-26;  
 Matches 48; Conservative 2; Mismatches 2; Indels 34; Gaps 1;  
 QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTR----- 30  
 DB 25 FVNHLCGSHLVEALYLVCGERGFFYTPKARREAEADLOCKDAELGEPAGGGLQPSALEA 84

QY 41 ----RGIVEQCCTSIQSLEYQENYCN 52  
 DB 85 PLQKRGIVEQCCTSIQSLEYQENYCN 110  
 RESULT 4  
 Q91BQ7 PRELIMINARY: PRT: 106 AA.  
 AC Q91BQ7;  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE Preproinsulin.  
 OS *Rana pipiens* (Northern leopard frog).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Amphibia; Batrachia; Anura; Neobatrachia; Ranidae; Rana.  
 OX NCRI\_TaxID:8404;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE:20362507; PubMed 10418274;  
 RA Irwin D.M., Sivarajah P.;  
 RT "Proinsulin cDNAs from the leopard frog, *Rana pipiens*: evolution of  
 RT proinsulin processing."  
 RL Comp. Biochem. Physiol. 125B:405-410(2000).  
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 DR EMBL: AF227187; AAF67285.1;  
 DR HSSP: P01315; 1SD8.  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1.  
 DR PRINTS: PK00277; INSULIN.  
 DR SMART: SM00078; IIGF; 1.  
 DR PROSITE: PS00262; INSULIN; 1.  
 SQ SEQUENCE 106 AA: 12183 MW: 3A870EFC70217F92 CRC64:

Query Match 74.7%; Score 219.5; DB 13; Length 106;  
 Best Local Similarity 49.4%; Pred. No. 1e-22;  
 Matches 41; Conservative 7; Mismatches 31; Gaps 1;  
 QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTR----- 31  
 DB 24 FQNGVLCGSHLVEALYMWCGRGFFYSRKRDLQPLVNHLCGSHLVEALYMWQVQSAFQKR 83  
 QY 42 -GIVEQCCTSIQSLEYQENYCN 52  
 DB 84 KPGIVEQCCHNTTSLYDLENYCN 106  
 RESULT 5  
 Q9RTA8 PRELIMINARY: PRT: 110 AA.  
 AC Q9RTA8;  
 DT 01-JUN-2001 (TRENBLrel. 17, Created)  
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE Preproinsulin.  
 OS *Pantodon buchholzi* (Butterflyfish).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;  
 CC Osteoglossiformes; Pantodontidae; Pantodon.  
 OX NCRI\_TaxID:8276;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE:21203577; PubMed 11304171;  
 RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;  
 RT "Molecular cloning of preproinsulin cDNAs from several  
 RT osteoglossomorphs and a cyprinid."  
 RL Mol. Cell. Endocrinol. 174:51-58(2001).  
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 DR EMBL: AF199588; AAK28712.1;  
 DR HSSP: P01308; 1NIS.

```

DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
SQ SEQUENCE 110 AA; 12324 MW; BDCCD659 472E06 CRC64;

Query Match
Best Local Similarity 68.5%; Score 201.5; DB 13; Length 110;
Matches 37; Conservative 8; Mismatches 5; Indels 35; Gaps 1;

QY 3 NHLLGSHLVVALYVCGERGFYTPK-----T 30
DB 26 SORLGGSHLVVALYVCGERGFYTPKROVDPLLLGFLPKSAQTEVAQFAFKDHAELRK 85
QY 31 ---RGIVEQCCTSGSLVQLENYCN 52
DB 86 KVRKGVVGGCHHPPNIFELQNYCN 108

RESULT 6
Q902N4
AC Q902N4 PRELIMINARY; PRT: 108 AA.
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 23, Last annotation update)
DE Insulin precursor
GN INS.
OS Brachydanio rerio (zebrafish) (Danio rerio);
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxId=7955;
RN [1]
RP MEDLINE:94425190; PubMed:10495291;
RA Argenton F., Zecchin E., Bortolussi M.;
RT "Early appearance of pancreatic hormone expressing cells in the
RT zebrafish embryo."
KL Mech. Dev. 87:217-221(1999).
CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL: AJ237750; CAC20109.1;
DR ZFIN: ZDB-GENE:980526-110; ins.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR00277; INSULIN.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
KW Signal.
FT SIGNAL. 1 24 POTENTIAL.
FT CHAIN. 24 53 INSULIN A CHAIN.
FT CHAIN. 86 108 INSULIN B CHAIN.
SQ SEQUENCE 108 AA; 11904 MW; 3105284E72AD6D25 CRC64;

Query Match
Best Local Similarity 66.5%; Score 195.5; DB 13; Length 108;
Matches 37; Conservative 5; Mismatches 7; Indels 43; Gaps 1;

QY 4 QHLLGSHLVVALYVCGERGFYTPK-----T 30
DB 27 QHLLGSHLVVALYVCGTFGFYTPKPKKSAQTEVAQFAFKDHAELRK 86
QY 31 RGIVEQCCTSGSLVQLENYCN 52
DB 87 RGIVEQCCHKPCSIFELQNYCN 108

RESULT 7
Q902N4
AC Q902N4 PRELIMINARY; PRT: 108 AA.
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 23, Last annotation update)
DE Insulin precursor
GN INS.
OS Brachydanio rerio (zebrafish) (Danio rerio);
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxId=7955;
RN [1]
RP MEDLINE:94425190; PubMed:10495291;
RA Argenton F., Zecchin E., Bortolussi M.;
RT "Early appearance of pancreatic hormone expressing cells in the
RT zebrafish embryo."
KL Mech. Dev. 87:217-221(1999).
CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL: AJ237750; CAC20109.1;
DR ZFIN: ZDB-GENE:980526-110; ins.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR00277; INSULIN.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
KW Signal.
FT SIGNAL. 1 24 POTENTIAL.
FT CHAIN. 24 53 INSULIN A CHAIN.
FT CHAIN. 86 108 INSULIN B CHAIN.
SQ SEQUENCE 108 AA; 11904 MW; 3105284E72AD6D25 CRC64;

Query Match
Best Local Similarity 66.5%; Score 195.5; DB 13; Length 111;
Matches 37; Conservative 5; Mismatches 7; Indels 43; Gaps 1;

QY 4 QHLLGSHLVVALYVCGERGFYTPK-----T 30
DB 27 QHLLGSHLVVALYVCGTFGFYTPKPKKSAQTEVAQFAFKDHAELRK 86
QY 31 RGIVEQCCTSGSLVQLENYCN 52
DB 87 RGIVEQCCHKPCSIFELQNYCN 108

RESULT 8
Q98TB0
AC Q98TB0 PRELIMINARY; PRT: 111 AA.
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 23, Last annotation update)
DE Preproinsulin (Fragment).
OS Chitala chitala (clown knifefish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;
OC Osteoglossiformes; Notopteridae; Chitala.
OX NCBI_TaxId=112153;
RN [1]
RP MEDLINE:21263577; PubMed:11361171;
RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;
RT "Molecular cloning of preproinsulin cDNAs from several
RT osteoglossomorphs and a cyprinid."
KL Mol. Cell. Endocrinol. 174:51-58(2001).
CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAM.IY.
DR EMBL: AF199586; AAK28710.1;
DR HSSP: P01308; 11PH.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR SMART: SM00078; IIGF; 1.
FT NON_TER. 111 111
SQ SEQUENCE 111 AA; 12483 MW; 247CA4431376329F CRC64;

Query Match
Best Local Similarity 66.3%; Score 195.5; DB 13; Length 111;
Matches 38; Conservative 3; Mismatches 9; Indels 36; Gaps 1;

```

```

DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 23, Last annotation update)
DE Preproinsulin.
OS Catla catla (catla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Catla.
OX NCBI_TaxId=72446;
RN [1]
RP SEQUENCE FROM N.A.
RA Bhattacharya S., Roy S.S., Dasgupta S., Ravikumar L., Mukherjee M.,
RA Bandyopadhyaya I., Wakabayashi K.;
RT "A new cell secreted insulin."
KL Submitted (APR-2001) to the EMBL/GenBank/JGI databases.
CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL: AF473623; AAK51468.1;
DR HSSP: P01908; 11NP.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR00277; INSULIN.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
SQ SEQUENCE 108 AA; 11881 MW; D713026E22EF5059 CRC64;

Query Match
Best Local Similarity 66.5%; Score 195.5; DB 13; Length 108;
Matches 37; Conservative 5; Mismatches 7; Indels 33; Gaps 1;

QY 4 QHLLGSHLVVALYVCGERGFYTPK-----T 30
DB 27 QHLLGSHLVVALYVCGTFGFYTPKPKKSAQTEVAQFAFKDHAELRK 86
QY 31 RGIVEQCCTSGSLVQLENYCN 52
DB 87 RGIVEQCCHKPCSIFELQNYCN 108

RESULT 8
Q98TB0
AC Q98TB0 PRELIMINARY; PRT: 111 AA.
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 23, Last annotation update)
DE Preproinsulin (Fragment).
OS Chitala chitala (clown knifefish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;
OC Osteoglossiformes; Notopteridae; Chitala.
OX NCBI_TaxId=112153;
RN [1]
RP MEDLINE:21263577; PubMed:11361171;
RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;
RT "Molecular cloning of preproinsulin cDNAs from several
RT osteoglossomorphs and a cyprinid."
KL Mol. Cell. Endocrinol. 174:51-58(2001).
CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAM.IY.
DR EMBL: AF199586; AAK28710.1;
DR HSSP: P01308; 11PH.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR SMART: SM00078; IIGF; 1.
FT NON_TER. 111 111
SQ SEQUENCE 111 AA; 12483 MW; 247CA4431376329F CRC64;

Query Match
Best Local Similarity 66.3%; Score 195.5; DB 13; Length 111;
Matches 38; Conservative 3; Mismatches 9; Indels 36; Gaps 1;

```

```

QY 3 NOHLCGSHLVEALYLVCGERGFFTYPTK ..... 29
DB 26 NOHLCGSHLVEALYLVCGERGFFTYPTKMKRDAEPLLGFLSPKSGLENEVDYFPKDGDL 85
QY 30 ---RGIVEQCCTISCSLYOLENYCN 52
DB 86 VMKRGIVEQCCHRCNIFDNOYCN 111

RESULT 9
ID Q98TA7 PRELIMINARY: PRT: 111 AA.
AC Q98TA7
DT 01-JUN-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Preproinsulin (Fragment)
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi
OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha
OC Osteoglossiformes; Hiodontidae; Hiodontidae
OX NCBI_TaxID:54904
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:21203577; PubMed:11306171;
RA Al Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;
RT "Molecular cloning of preproinsulin cDNAs from several
osteoglossomorphs and a cyprinid."
RL Mol. Cell. Endocrinol. 174:51-58(2001).
CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
DR EMBL: AF282408; AAK54684.1;
DR HSSP: P01308; IHP.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
FT NON_TER 111
SQ SEQUENCE 110 AA: 12343 MW: 80800770452606 CRC64:

Query Match 65.8%; Score 194.5; DB 13; Length 110;
Best Local Similarity 42.4%; Pred. No. 5.8e-18;
Matches 36; Conservative 7; Mismatches 7; Indels 45; Gaps 1;

QY 3 NOHLCGSHLVEALYLVCGERGFFTYPTK ..... 30
DB 26 NOHLCGSHLVEALYLVCGERGFFTYPTKMKRDAEPLLGFLSPKSGLENEVDYFPKDGDL 85
QY 31 ---RGIVEQCCTISCSLYOLENYCN 52
DB 86 VMKRGIVEQCCHRCNIFDNOYCN 111

RESULT 13
Q98TA7 PRELIMINARY: PRT: 111 AA.
AC Q98TA7
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 20, Last annotation update)
DE Preproinsulin (Fragment)
OS Osteoglossum bicirrhosum (silver arowana)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha
OC Osteoglossiformes; Osteoglossidae; Osteoglossum
OX NCBI_TaxID:109271;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:21203577; PubMed:11306171;
RA Al Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;
RT "Molecular cloning of preproinsulin cDNAs from several
osteoglossomorphs and a cyprinid."

```

```

RL Mol. Cell. Endocrinol. 174:51-58(2001).
CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
DR EMBL: AF199589; AAK28711.1;
DR HSSP: P01315; IHP.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR00277; INSULIN.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
FT NON_TER 111
SQ SEQUENCE 111 AA: 12491 MW: 8093190204866823 CRC64:

Query Match 65.1%; Score 191.5; DB 13; Length 111;
Best Local Similarity 41.2%; Pred. No. 2.1e-19;
Matches 35; Conservative 10; Mismatches 5; Indels 35; Gaps 1;

QY 4 NOHLCGSHLVEALYLVCGERGFFTYPTK ..... 30
DB 26 NOHLCGSHLVEALYLVCGERGFFTYPTKMKRDAEPLLGFLSPKSGLENEVDYFPKDGDL 86
QY 31 ---RGIVEQCCTISCSLYOLENYCN 52
DB 87 VMKRGIVEQCCHRCNIFDNOYCN 111

RESULT 11
Q98TA7 PRELIMINARY: PRT: 67 AA.
AC Q98TA7
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Preproinsulin (Fragment)
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha
OC Osteoglossiformes; Mormyridae; Gnathionemus
OX NCBI_TaxID:42645;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:21203577; PubMed:11306171;
RA Al Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;
RT "Molecular cloning of preproinsulin cDNAs from several
osteoglossomorphs and a cyprinid."
RL Mol. Cell. Endocrinol. 174:51-58(2001).
CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
DR EMBL: AF199587; AAK28711.1;
DR HSSP: P01308; IHP.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
FT NON_TER 87
SQ SEQUENCE 87 AA: 8774 MW: 80448050245355 CRC64:

Query Match 63.8%; Score 187.5; DB 13; Length 87;
Best Local Similarity 42.9%; Pred. No. 5.8e-18;
Matches 36; Conservative 5; Mismatches 8; Indels 35; Gaps 1;

QY 4 OHLCGSHLVEALYLVCGERGFFTYPTK ..... 30
DB 4 OHLCGSHLVEALYLVCGERGFFTYPTKMKRDAEPLLGFLSPKSGLENEVDYFPKDGDL 64
QY 31 ---RGIVEQCCTISCSLYOLENYCN 52
DB 64 VMKRGIVEQCCHRCNIFDNOYCN 87

RESULT 12
Q98TB1

```

```

ID Q98TB1 PRELIMINARY: PRT: 108 A.
AC Q98TB1:
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Preproinsulin (Fragment).
OS Catostomus commersoni (White sucker).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Catostomidae; Catostomus.
OX NCBI_TaxID=7971;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 21203577; PubMed: 11306171;
RA Al-Mabrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;
RT "Molecular cloning of preproinsulin cDNAs from several
RT osteoglossomorphs and a cyprinid."
RL Mol. Cell. Endocrinol. 174:51-56 (2000).
CC -1. SIMILARITY: BELONGS TO THE INSULIN/F/RELAXIN FAMILY.
DE EMBL: AF194585; AAK28709.1;
DR HSSP: P01308; 11PH.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
FT NON_TER 108 108
SQ SEQUENCE 108 AA: 1187; MW: 1426310696; FSAFC8 CRC64;

Query Match: 63.4%; Score 186.5; DB 13; Length 108;
Best Local Similarity 43.9%; Pred. No. 1,1e-17;
Matches 36; Conservative 4; Mismatches 9; Indels 33; Gaps 1;

QY 4 QHLCGSHLVVLCVCGRGFFYTPK .....T 30
DB 27 QHLCGSHLVVLCVCGRGFFYTPK .....T 30

QY 31 NCIVEQCCTSCSLYULENYCN 52
DB 87 NCIVEQCCTSCSLYULENYCN 108

RESULT 13
Q98TB2 PRELIMINARY: PRT: 91 AA.
AC Q98TB2:
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Preproinsulin (Fragment).
OS Ambloplites rupestris (Rock bass).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoleiostei;
OC Centrarchidae; Acanthopterygii; Perciformes; Percoidae;
OX NCBI_TaxID=109273;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 21203577; PubMed: 11306171;
RA Al-Mabrouki A.A., Irwin D.M., Youson J.H.;
RT "Molecular cloning of preproinsulin cDNA from the rock bass."
RL Submitted (01-1999) to the EMBL/GenBank/DBJ databases.
CC -1. SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -1. SIMILARITY: BELONGS TO THE INSULIN/F/RELAXIN FAMILY.
DE EMBL: AF194584; AAK28708.1;
DR HSSP: P01308; 11PH.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
FT NON_TER 91 91
SQ SEQUENCE 91 AA: 10100 MW: 1106682560; FSAFC9 CRC64;

```

```

Query Match: 63.18%; Score 185.5; DB 13; Length 91;
Best Local Similarity 40.9%; Pred. No. 1,1e-17;
Matches 36; Conservative 5; Mismatches 8; Indels 39; Gaps 1;

QY 4 QHLCGSHLVVLCVCGRGFFYTPK .....T 29
DB 4 QHLCGSHLVVLCVCGRGFFYTPK .....T 29

QY 30 ---- TCGIVEQCCTSCSLYULENYCN 52
DB 64 MEMVKRGIVCGCCHHPCNIFDLRYCN 91

RESULT 14
Q98Z81 PRELIMINARY: PRT: 65 AA.
AC Q98Z81:
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Insulin (Fragment).
OS Gorilla gorilla (gorilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Gorilla.
OX NCBI_TaxID=9593;
RN [1]
RP SEQUENCE FROM N.A.
RA O'Huigin C., Tichy H., Klein J.;
RT "Molecular evolution in higher primates: gene specific and organism
RT specific characteristics."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY092023; AAM76640.1;
FT NON_TER 1 1
FT NON_TER 65 65
SQ SEQUENCE 65 AA: 5920 MW: 8772017FD8HCAHEA CRC64;

Query Match: 49.7%; Score 146; DB 6; Length 65;
Best Local Similarity 47.7%; Pred. No. 2e-12;
Matches 31; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

QY 7 CGSHLVVLCVCGRGFFYTPK .....T 32
DB 3 CGSHLVVLCVCGRGFFYTPK .....T 32

QY 53 IVEQC 37
DB 61 IVEQC 65

RESULT 15
Q98Z80 PRELIMINARY: PRT: 65 AA.
AC Q98Z80:
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Insulin (Fragment).
OS Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pongo.
OX NCBI_TaxID=9600;
RN [1]
RP SEQUENCE FROM N.A.
RA O'Huigin C., Tichy H., Klein J.;
RT "Molecular evolution in higher primates: gene specific and organism
RT specific characteristics."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY092024; AAM76641.1;
FT NON_TER 1 1
FT NON_TER 65 65
SQ SEQUENCE 65 AA: 5920 MW: 8772017FD8HCAHEA CRC64;

Query Match: 49.7%; Score 146; DB 6; Length 65;

```

Best Local Similarity 47.7%; Pred. No. 2e-12;  
Matches 31; Conservative 0; Mismatches 0; Indels 34; Gaps 1;  
QY 7 CGSHLVEALYLCGERGFFYTPKT-----RG 32  
Db 1 CGSHLVEALYLCGERGFFYTPKT-----II 11  
QY 33 IVFQC 37  
Db 61 IVFQC 65

Search completed: September 15, 2003, 12:03:30  
Job time : 21.3154 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2003, 12:31:20 ; Search time 31.642 seconds  
(without alignments)  
536.746 Million cell updates/sec

Title: us-09-423-100-6

Perfect score: 587

Sequence: 1 MFPTPLSRFLDNALRAHR.....IVEOCTSTCSLYOLENYCN 107

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 res.jues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	587	100.0	107	20	hGH-mini-proinsulin
2	555.5	94.6	150	20	Chimeric protein,
3	315.5	53.7	116	17	SOD-proinsulin hyb
4	304	51.8	63	15	Human pro-insulin
5	304	51.8	117	17	hGH-mini-proinsulin
6	302.5	51.5	137	16	Mating factor alph
7	299	50.9	56	15	Human pro-insulin
8	299	50.9	56	16	AA198665
9	299	50.9	96	15	AA198899

10	299	50.9	96	16	AA198662
11	299	50.9	145	16	AA198694
12	299	50.9	146	16	AA198695
13	294	50.1	52	20	AA198699
14	293	49.9	57	11	AA198700
15	288.5	49.1	160	16	AA198705
16	287	48.9	52	12	AA198706
17	287	48.9	65	19	AA198707
18	287	48.9	138	1	AA198708
19	284.5	48.5	58	1	AA198709
20	284.5	48.5	59	1	AA198710
21	284.5	48.5	65	1	AA198711
22	284.5	48.5	109	10	AA198712
23	284.5	48.5	123	18	AA198713
24	284.5	48.5	123	19	AA198714
25	284.5	48.5	124	19	AA198715
26	284.5	48.5	124	24	AA198716
27	284.5	48.5	124	24	AA198717
28	284.5	48.5	125	18	AA198718
29	284	48.4	138	10	AA198719
30	284	48.4	138	17	AA198720
31	284	48.4	140	16	AA198721
32	284	48.4	140	16	AA198722
33	283.5	48.3	53	1	AA198723
34	283.5	48.3	53	1	AA198724
35	283.5	48.3	117	11	AA198725
36	283.5	48.3	408	22	AA198726
37	282	48.0	94	10	AA198727
38	282	48.0	120	18	AA198728
39	281.5	48.0	60	3	AA198729
40	281.5	48.0	105	1	AA198730
41	281.5	48.0	153	21	AA198731
42	281	47.9	102	10	AA198732
43	281	47.9	104	16	AA198733
44	280.5	47.8	55	8	AA198734
45	280.5	47.8	89	17	AA198735

#### ALIGNMENTS

#### RESULT 1

AY42860 ID AY42860 standard; protein; 107 AA.

XX AC AY42860;

DT 19-JAN-2000 (first entry)

DE hGH-mini-proinsulin chimeric protein.

XX Insulin; precursor; growth hormone; chaperone; intramolecular;  
KW folding; conformation; chimeric protein; cleavable; recombinant;  
KW production; yield.

XX OS Synthetic.

XX OS Homo sapiens.

PD 07-OCT-1999.

XX PF 31-MAR-1998; 98WO-CN00052.

XX PR 31-MAR-1998; 98WO-CN00052.

XX (TONG-) TONGHUA GANTECH BIOTECHNOLOGY LTD.

XX WPI; 1999-610839/52.

PT New chimeric proteins containing human growth hormone fragment, used

Fusion protein con  
Mating factor alph  
Human insulin prec  
Proinsulin analogu  
Glycosylphosphatid  
Example of human i  
Preproinsulin 1.  
pK142 modified in  
Modified pJ859 mod  
Modified pJ859 mod  
N-terminally exten  
Amino acids encode  
EEAEPK-M13 insulin  
DNA construct pAK7  
pAK855 protein seq  
Synthetic leader p  
Insulin precursor  
EEAEPK-M15 insulin  
Amino acids encode  
S. cerevisiae MF a  
Mating factor alph  
Mating factor alph  
Di-Arg-(B31-32)-Hu  
Ins1 double-chain  
Protein sequence o  
A Bacillus pectate  
Amino acids encode  
EEAEPK-M11 insulin  
Human proinsulin a  
Amino acids encode  
Human proinsulin  
Signal-leader-p-in  
Yeast signal/leade  
Sequence encoded b  
Signal peptide/lea

PT particularly for the production of human insulin -  
 PS Claim 13; Page 30; 46pp; English.  
 XX  
 CC This sequence represents a chimeric protein, hGH-mini-proinsulin.  
 CC This chimeric protein contains an N-terminal fragment of human growth  
 CC hormone (hGH) of the sequence given in AAY42855, a cleavable peptide  
 CC linker (AAY42857), and a human insulin precursor comprising insulin  
 CC A and B chains (AAY42859). The hGH portion of the chimeric protein acts  
 CC as an intramolecular chaperone (IMC) for the insulin precursor,  
 CC enabling it to fold correctly. The cleavable peptide linker has a  
 CC C-terminal Arg residue which enables the hGH portion of the  
 CC chimeric protein to be removed after folding has taken place. Production  
 CC of recombinant human insulin via an hGH-proinsulin chimeric protein can  
 CC provide human insulin with correctly linked cysteine bridges with  
 CC fewer necessary procedural steps, and hence resulting in a higher yield  
 CC of human insulin. The IMC sequences not only protect insulin sequences  
 CC from intracellular degradation by a microorganism host, but also promote  
 CC the folding of the fused insulin precursor, facilitate the solubility of  
 CC the fusion protein and decrease the intermolecular interactions among  
 CC the fusion proteins, thus allowing folding of the fused insulin precursor  
 CC at commercially useful high concentrations. The procedural steps of  
 CC cyanogen bromide cleavage, oxidative sulphitolysis and related  
 CC purification steps can thus be eliminated, along with the use of high  
 CC concentrations of mercaptan or the use of hydrophobic absorbent resins.  
 XX  
 SQ Sequence 107 AA;  
 Query Match 100.0%; Score 587; DB 20; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 4e-42;  
 Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MEPTIPLSRLEFDNMLRAHRLHQLAFDTYQEFEEAYIPKEOKYSFLQNPGLTGRPFVNHQ 60  
 DB 1 MEPTIPLSRLEFDNMLRAHRLHQLAFDTYQEFEEAYIPKEOKYSFLQNPGLTGRPFVNHQ 60  
 QY 61 LCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYQLENYCN 107  
 DB 61 LCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYQLENYCN 107  
 RESULT 2  
 AAY42861  
 ID AAY42861 standard; protein: 150 AA.  
 AC AAY42861;  
 DT 19-JAN-2000 (first entry)  
 DE Chimeric protein, SEQ ID 7.  
 XX  
 KW Insulin; precursor: growth hormone; chaperone: intramolecular;  
 KW folding; conformation: chimeric protein; cleavable; recombinant;  
 KW production; yield.  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 PN W09950302-A1.  
 XX  
 PD 07-OCT-1999.  
 PF 31-MAR-1998; 98WO-CN00052.  
 XX  
 PR 31-MAR-1998; 98WO-CN00052.  
 XX  
 PA (TONG-) TONGHUA GANTECH BIOTECHNOLOGY LTD.  
 XX  
 PI Gan Z;  
 XX  
 DR WPI: 1999-610839/52.  
 XX  
 PT New chimeric proteins containing human growth hormone fragment, used

PT particularly for the production of human insulin -  
 PS Claim 14; Page 30-31; 46pp; English.  
 XX  
 CC This sequence represents a chimeric protein, which contains an  
 CC N-terminal fragment of human growth hormone (hGH) of the sequence given  
 CC in AAY42856, a cleavable peptide linker (AAY42857), and a human insulin  
 CC precursor comprising insulin A and B chains (AAY42859). The hGH portion  
 CC of the chimeric protein acts as an intramolecular chaperone (IMC) for  
 CC the insulin precursor, enabling it to fold correctly. The cleavable  
 CC peptide linker has a C-terminal Arg residue which enables the hGH portion  
 CC of the chimeric protein to be removed after folding has taken place.  
 CC Production of recombinant human insulin via an hGH-proinsulin chimeric  
 CC protein can provide human insulin with correctly linked cysteine bridges  
 CC with fewer necessary procedural steps, and hence resulting in a higher  
 CC yield of human insulin. The IMC sequences not only protect insulin  
 CC sequences from intracellular degradation by a microorganism host, but  
 CC also promote the folding of the fused insulin precursor, facilitate the  
 CC solubility of the fusion protein and decrease the intermolecular  
 CC interactions among the fusion proteins, thus allowing folding of the  
 CC fused insulin precursor at commercially useful high concentrations. The  
 CC procedural steps of cyanogen bromide cleavage, oxidative sulphitolysis  
 CC and related purification steps can thus be eliminated, along with the use  
 CC of high concentrations of mercaptan or the use of hydrophobic absorbent  
 CC resins.  
 XX  
 SQ Sequence 150 AA;  
 Query Match 94.6%; Score 555.5; DB 20; Length 150;  
 Best Local Similarity 71.3%; Pred. No. 2.4e-39;  
 Matches 107; Conservative 0; Mismatches 0; Indels 43; Gaps 1;  
 QY 1 MEPTIPLSRLEFDNMLRAHRLHQLAFDTYQEFEEAYIPKEOKYSFLQNP----- 49  
 DB 1 MEPTIPLSRLEFDNMLRAHRLHQLAFDTYQEFEEAYIPKEOKYSFLQNPOTLSFSFSEIP 60  
 QY 50 -----LGTGPRFVNHQGLSHLVEALYLVCGER 77  
 DB 61 TPSNREETQOKSNLELLRLISLLLIQSNLEPVLQGTGPRFVNHQGLSHLVEALYLVCGER 120  
 QY 78 GFFYTPKTRGIVEQCCTSCISLYQLENYCN 107  
 DB 121 GFFYTPKTRGIVEQCCTSCISLYQLENYCN 150  
 RESULT 3  
 AAR98897  
 ID AAR98897 standard; Protein: 116 AA.  
 AC AAR98897;  
 DT 03-FEB-1997 (first entry)  
 DE SOD-proinsulin hybrid polypeptide.  
 KW Insulin; proinsulin; hybrid polypeptide; protein folding;  
 KW enzymatic cleavage; cyanogen bromide; sulphitolysis.  
 OS Homo sapiens.  
 XX  
 PN W09620724-A1.  
 XX  
 PD 11-JUL-1996.  
 PF 29-DEC-1994; 94WO-US13268.  
 XX  
 PR 29-DEC-1994; 94WO-US13268.  
 XX  
 PR 10-JAN-1995; 95ZA-0000142.  
 XX  
 PA (BIOT-) BIO-TECHNOLOGY GENERAL CORP.  
 XX  
 PI Gorecki M, Hartman JR, Mendelovitz S;  
 XX



4:

KTG 87  
1111  
KTG 117

८८

```

XX  Sequence 56 AA:
SQ
  Query Match 50.9%; Score 299; DB 15; Length 56;
  Best Local Similarity 100.0%; Pred. No. 3.2e-18;
  Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 55 RFVNHLCGSHLVEALYLIVCGGEGFFTPKTRGIVEQCCTSIQSLYOLENYCN 107
DB 4 RFVNHLCGSHLVEALYLIVCGGEGFFTPKTRGIVEQCCTSIQSLYOLENYCN 56

RESULT 2
AAR78665
ID AAR78665 standard; protein; 56 AA.
AC AAR78665;
XX
XX 01-APR-1996 (first entry)
DE Proinsulin sequence 1.
KW Proinsulin; post-translational modification; recombinant production;
KW protein folding; conformation.
XX
OS Synthetic.
XX
XX Key Location/Qualifiers
FH Region 1..4
FT /label= R2
FT /note= "a peptide of 4 amino acids"
FT Peptide 5..34
FT /label= R1-(B2-B29); Y
FT Region 35
FT /note= "human insulin B-chain"
FT Peptide 36..56
FT /label= Gly-(A2-A20)-R3
FT /note= "human insulin A-chain"
XX
XX EP668292-A2.
XX
XX 23-AUG-1995.
XX
XX 09-FEB-1995; 95EP-0101748.
XX
XX 18-FEB-1994; 94DE-4405179.
XX (FARH ) HOECHST AG.
XX
XX Gerl M, Ludwig J, Obermeier R, Sabel W;
XX WPI; 1995-284754/38.
XX
XX Isolation of insulin that is correctly post-translationally
XX processed - by reacting proinsulin with a mercaptan in the presence
XX of a chaotropic agent and purification, after absorption to hydrophobic
XX resin
XX
XX Example 2; Page 13; 16pp; German.
XX
XX The present sequence is an example of a proinsulin molecule corresp.
XX to the general formula R2-R1-(B2-B29)-Y-X-Gly-(A2-A20)-R3 (I); in
XX formula (II), X = Lys, Arg or a peptide of 2-35 amino acids contg.
XX Lys or Arg at the N- and C-termini; Y = a natural amino acid; R1 =
XX Phe or a bond; R2 = H, Arg, Lys, a peptide of 2-45 amino acids
XX contg. Arg or Lys at the N- and C-termini; R3 = a natural amino
XX acid; (A2-A20) and (B2-B29) are the insulin A- and B-chain sequences
XX from human or other insulin. The proinsulin molecule (produced in
XX recombinant E.coli) is reacted with mercaptan at a ratio of 2-10 SH
XX residues of mercaptan per Cys residue of proinsulin. The reaction
XX takes place in the presence of a chaotropic auxiliary agent at
XX pH 10-11 and results in proinsulin with correctly linked cystine

```

---

```

CC bridges. Reaction with trypsin and opt. carboxypeptidase B yields
CC correctly folded insulin. The insulin is isolated by absorption on
CC a hydrophobic resin.
XX
XX Sequence 56 AA;
SQ
  Query Match 50.9%; Score 299; DB 16; Length 56;
  Best Local Similarity 100.0%; Pred. No. 3.2e-18;
  Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 55 RFVNHLCGSHLVEALYLIVCGGEGFFTPKTRGIVEQCCTSIQSLYOLENYCN 107
DB 4 RFVNHLCGSHLVEALYLIVCGGEGFFTPKTRGIVEQCCTSIQSLYOLENYCN 56

RESULT 9
AAR68899
ID AAR68899 standard; peptide; 96 AA.
XX
XX AC AAR68899;
XX
XX 25-MAR-2003 (updated);
DT 02-MAR-1995 (first entry);
XX
XX DE Human pro-insulin 2.
XX
XX KW Pro-insulin; A chain; B chain; C-chain; disulphide;
KW mercaptan; chaotropic agent.
XX
XX OS Homo sapiens.
XX
XX PN EP600372-A1.
XX
XX 08-JUN-1994.
XX
XX 25-NOV-1993; 93EP-0118993.
XX
XX 02-DEC-1992; 92DE-4240420.
XX (FARH ) HOECHST AG.
XX
XX Gerl M, Ludwig J, Obermeier R, Sabel W;
XX WPI; 1994-177718/22.
XX
XX Prodn. of pro-insulin with correct disulphide bridges by
XX treating recombinant precursor protein with mercaptan in alkali;
XX and in presence of chaotropic agent, then isolation on
XX hydrophobic resin
XX
XX Disclosure; Page 11; 15pp; German.
XX
XX Pro insulin is produced by treating recombinant precursor protein
XX with a mercaptan to produce 2-10 SH residues per Cys residue, in
XX presence of a chaotropic agent and in an aqueous medium of pH 10-11;
XX treating the prod. with 3-50 g hydrophobic adsorber resin per l aq.
XX medium of pH 4-7, isolating the adsorbed resin and pro-insulin and
XX desorbing the pro-insulin. This method produces pro-insulin with
XX correctly bonded Cys bridges. Compared with known methods it
XX involves fewer stages (esp. no sulphonylation or cyanogen bromide
XX cleavage) and overall losses during purification are reduced, i.e.
XX the process is quicker and gives better yields.
XX Sequences of insulin chain A, B and C are given in AAR68895-97.
XX Sequences of pro-insulin 1-4 are given in AAR68898-901.
XX (Updated on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 96 AA;
SQ
  Query Match 50.9%; Score 299; DB 15; Length 96;
  Best Local Similarity 100.0%; Pred. No. 5.1e-18;
  Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 55 RFVNHLCGSHLVEALYLIVCGGEGFFTPKTRGIVEQCCTSIQSLYOLENYCN 107

```

Db 44 RFVNHLCGSHLVEALVLCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 96  
 |||

## RESULT 10

AAR78662  
 ID AAR78662 standard; protein: 96 AA.

XX AAR78662;

AC 03-APR-1996 (first entry)

XX Fusion protein contg. proinsulin sequence 3.

XX Proinsulin: post-translational modification; recombinant production;  
 XX protein folding; conformation.

XX Synthetic.

XX Key Location/Qualifiers  
 FH Region 41..44  
 FT /label= R2  
 FT /note= "a peptide of 4 amino acids"  
 FT 45..74  
 FT /label= R1-(B2-B29)-Y  
 FT /note= "human insulin B-chain"  
 FT 75  
 FT /label= X  
 FT 76..96  
 FT /label= Gly-(A2-A20)-R3  
 FT /note= "human insulin A-chain"

XX EP668292-A2.

XX 23-AUG-1995.

XX 09-FEB-1995; 95EP-0101748.

XX 18-FEB-1994; 94DE-4405179.

XX (FAKH) HOECHST AG.

XX Gerl M, Ludwig J, Obermeier R, Sabel W;

XX WPI: 1995-284754/38.

XX Isolation of insulin that is correctly post-translationally  
 PT processed - by reacting pro-insulin with a mercaptan in the presence  
 PT of a chaotropic agent and purification after absorption to hydrophobic  
 PT resin

XX Example 2; Page 8; 16pp; German.

XX The present sequence is that of a fusion protein, produced in E.coli  
 CC which contains an example of a proinsulin molecule corresp.  
 CC to the general formula R2-R1-(B2-B29)-Y-X-Gly-(A2-A20)-R3 (II). In  
 CC formula (II), X = Lys, Arg or a peptide of 2-35 amino acids contg.  
 CC Lys or Arg at the N- and C-termini; Y = a natural amino acid; R1 =  
 CC Phe or a bond; R2 = H, Arg, Lys, a peptide of 2-45 amino acids  
 CC contg. Arg or Lys at the N- and C-termini; R3 = a natural amino  
 CC acid; (A2-A20) and (B2-B29) are the insulin A- and B-chain sequences  
 CC from human or other insulin. The proinsulin molecule, released by  
 CC cyanogen bromide, is reacted with mercaptan at a ratio of 2-10 SH  
 CC residues of mercaptan per Cys residue of proinsulin. The reaction  
 CC takes place in the presence of a chaotropic auxiliary agent at  
 CC pH 10-11 and results in proinsulin with correctly linked cystine  
 CC bridges. Reaction with trypsin and opt. carboxypeptidase B yields  
 CC correctly folded insulin. The insulin is isolated by absorption on  
 CC a hydrophobic resin.

XX Sequence 96 AA;

Query Match 50.9%; Score 299; DB 16; Length 96;

Best Local Similarity 100.0%; Pred. No. 5.le-18;  
 Matches 53; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 55 RFVNHLCGSHLVEALVLCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 107  
 |||

DB 44 RFVNHLCGSHLVEALVLCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 96

## RESULT 11

AAR71694  
 ID AAR71694 standard; Protein: 115 AA.

XX AAR71694;

XX 25-MAR-2003 (updated)

DT 20-NOV-1995 (first entry)

XX Mating factor alpha 1-Insulin precursor ArgB1, ArgB31 N-terminal.

XX Human insulin precursor ArgB31; diabetes; Zinc ion complex;  
 XX mating factor alpha 1; N-terminal EEAEEAR.

XX Homo sapiens.

XX Key Location/Qualifiers  
 FH Protein 1..85  
 FT /label= mating factor alpha-1  
 FT 86..93  
 FT /label= N-terminal peptide  
 FT 94..124  
 FT /label= B-chain  
 FT 125..145  
 FT /label= A-chain

XX W09507931-A1.

XX 23-MAR-1995.

XX 16-SEP-1994; 94WO-DK00347.

XX 17-SEP-1993; 93DK-00C1044.

XX 02-FEB-1994; 94US-0190829.

XX (NOVO) NOVO-NORDISK AS.

XX Andersen AS, Halstrom JB, Avelund S, Jonassen I;

XX Markussen J;

XX WPI: 1995-131314/17.

XX N-PSDB; AAQ86429.

XX Acylated insulin deriv. which may be present as a Zinc ion  
 PT complex - is used to treat diabetes and is rapid acting.

XX Example 5; Pages 82-83; 100pp; English.

XX AAQ86429 encodes AAR71694 mating factor alpha 1-Insulin precursor  
 CC ArgB1, ArgB31 N-terminal EEAEEAR. The insulin precursor comprises  
 CC the B and A chains of a claimed human insulin derivative preceded  
 CC by the N-terminal amino acids EEAEEAR. In the final claimed compsn.  
 CC they are covalently connected via disulphide bonds between Cys  
 CC residues A7/B7 and A20/B19. The derivative, which may be present  
 CC as a zinc ion complex, can be used as a fast action treatment for  
 CC diabetes.  
 CC (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 145 AA;

Query Match 50.9%; Score 299; DB 16; Length 145;

Best Local Similarity 100.0%; Pred. No. 7.3e-18;  
 Matches 53; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 55 RFVNHLCGSHLVEALVLCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 107

Db 93 RFVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYQLENYCN 145  
|||||  
RESULT 12  
AAR71695  
ID AAR71695 standard; Protein: 146 AA.  
XX  
AC AAR71695;  
XX  
XX 25-MAR-2003 (updated)  
DT 20-NOV-1995 (first entry)  
XX  
XX Mating factor alpha 1-Insulin precursor ArgB1, ArgB31 N-terminal.  
DE Human insulin precursor ArgB1, ArgB31; diabetes; zinc ion complex;  
KW mating factor alpha 1; N-terminal EEAEAEAE.  
XX  
XX Homo sapiens.  
OS  
FH Key Location/Qualifiers  
FT Protein 1..85  
FT /label= mating factor alpha-1  
FT Peptide 86..94  
FT /label= N-terminal peptide  
FT Peptide 95..125  
FT /label= B-chain  
FT Peptide 126..146  
FT /label= A-chain  
XX  
PN W09507931-AL.  
XX  
XX 23-MAR-1995.  
XX  
XX 16-SEP-1994; 94WO-DK00347.  
XX  
XX 17-SEP-1993; 93DK-0001044.  
XX 02-FEB-1994; 94US-0190829.  
XX (NOVO ) NOVO-NORDISK AS.  
XX  
XX Andersen AS, Halstrom JB, Havelund S, Jonassen I;  
PI Markussen J;  
XX  
XX WPI: 1995-131314/17.  
XX N-PSDB: AAQ86432.  
XX  
XX Acylated insulin deriv. which may be present as a zinc ion  
PT complex - is used to treat diabetes and is rapid acting.  
XX  
XX Example 6; Page 85; 100pp; English.  
XX  
XX AAQ86432 encodes AAR71695 mating factor alpha 1-Insulin precursor  
CC ArgB1, ArgB31 N-terminal EEAEAEAE. The insulin precursor comprises  
CC the B and A chains of a claimed human insulin derivative preceded  
CC by the N-terminal amino acids EEAEAEAE. In the final claimed compn.  
CC residues A7/B7 and A20/B19. The derivative, which may be present  
CC as a zinc ion complex, can be used as a fast action treatment for  
CC diabetes.  
XX  
XX (Updated on 25-MAR-2003 to correct PN field.)  
XX  
XX Sequence 146 AA;  
SQ  
Query Match 50.9%; Score 299; DB 16; Length 146;  
Best Local Similarity 100.0%; Pred. No. 7.4e-18;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 55 RFVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYQLENYCN 107  
|||||  
Db 94 RFVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYQLENYCN 146  
|||||

RESULT 13  
AAY42859  
ID AAY42859 standard; protein: 52 AA.  
XX  
AC AAY42859;  
XX  
XX 19-JAN-2000 (first entry)  
DT  
XX  
XX Human insulin precursor, SFO ID 5.  
DE  
XX  
XX Insulin; precursor; growth hormone; chaperone; intramolecular;  
KW folding; conformation; chimeric protein; cleavable; recombinant;  
KW production; yield.  
XX  
XX Homo sapiens.  
OS  
PN W09950302-AL.  
XX  
XX 07-OCT-1999.  
XX  
XX 31-MAR-1998; 98WO-CN00052.  
XX  
XX 31-MAR-1998; 98WO-CN00052.  
XX (TONG-) TONGHUA GANTECH BIOTECHNOLOGY LTD.  
XX Gan Z;  
XX  
XX WPI: 1999-610839/52.  
XX  
XX New chimeric proteins containing human growth hormone fragment, used  
PT particularly for the production of human insulin  
XX  
XX Claim 12; Page 29-30; 46pp; English.  
XX  
XX This sequence represents a human insulin precursor comprising insulin  
CC A and B chains. This insulin precursor is a component of the chimeric  
CC proteins hGH-mini-proinsulin (AAY42860) and the chimeric protein  
CC given in AAY42861. These chimeric proteins additionally contain an  
CC N-terminal fragment of human growth hormone (hGH) and a cleavable  
CC peptide linker (AAY42857). The hGH portion of the chimeric protein acts  
CC as an intramolecular chaperone (IMC) for the insulin precursor,  
CC enabling it to fold correctly. The cleavable peptide linker has a  
CC C-terminal Arg residue which enables the hGH portion of the  
CC chimeric protein to be removed after folding has taken place. Production  
CC of recombinant human insulin via an hGH-proinsulin chimeric protein can  
CC provide human insulin with correctly linked cysteine bridges with  
CC fewer necessary procedural steps, and hence resulting in a higher yield  
CC of human insulin. The IMC sequences not only protect insulin sequences  
CC from intracellular degradation by a microorganism host, but also promote  
CC the folding of the fused insulin precursor, facilitate the solubility of  
CC the fusion proteins and decrease the intermolecular interactions among  
CC the fusion proteins, thus allowing folding of the fused insulin precursor  
CC at commercially useful high concentrations. The procedural steps of  
CC cyanogen bromide cleavage, oxidative sulphytolysis and related  
CC purification steps can thus be eliminated, along with the use of high  
CC concentrations of mercaptan or the use of hydrophobic absorbent resins.  
XX  
XX Sequence 52 AA;  
SQ  
Query Match 50.1%; Score 294; DB 20; Length 52;  
Best Local Similarity 100.0%; Pred. No. 7.9e-18;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 56 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYQLENYCN 107  
|||||  
Db 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYQLENYCN 52  
|||||  
RESULT 14  
AAR04582  
ID AAR04582 standard; protein: 57 AA.  
XX

```

AC AAR04582;
XX
XX 25-MAR-2003 (updated)
DT 14-SEP-1990 (first entry)
XX
XX Proinsulin analogue with a Lys residue linking the A and B chains.
DE
XX insulin fusion protein; pro-insulin analogue; tandemistate;
KW Lys-Lys bridge; ds
XX
XX synthetic.
OS
XX
XX
FH Key Location/Qualifiers
FT misc-difference 36
FT Peptide /label=Lys residue linking insulin B chain to A chain
FT Peptide 1..35
FT Peptide /label= Insulin B chain
FT Peptide 37..57
FT Peptide /label= Insulin A chain
XX
XX EP367163-A.
PN
XX
XX 09-MAY-1990.
PD
XX
XX 28-OCT-19 9; 89EP-0120056.
XX
XX 03-NOV-1988; 88DE-3837273.
PR 19-AUG-1989; 89DE-3927449.
XX
XX (FARRH) HOECHST AG.
PA
XX
XX Koller KP, Riess GJ, Uhlmann E, Wallmeier H;
PI
XX
XX WPI; 1990-141149/19.
DR
XX
XX N-PSDB; AAQ04335.
XX
XX New insulin fusion proteins comprise pro-insulin analogue linked
PT to tandemistate
XX
XX
XX Disclosure; Page ?; 7pp; German.
XX
XX This sequence is joined to the C-terminus of an N-terminal fragment
CC comprising opt. modified tandemistate. This fusion protein
CC may be converted into human insulin using known methods. The synthetic
CC gene was prepared by the phosphoramidite method.
CC
CC See also AAQ04336.
CC
CC (Updated on 25-MAR-2003 to correct PR field.)
CC
CC (Updated on 25-MAR-2003 to correct PI field.)
XX
XX Sequence 57 AA;
SQ
Query Match 49.9%; Score 293; DB 11; Length 57;
Best Local Similarity 96.2%; Pred. No. 1e-17;
Matches 51; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 55 RFVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYQLENYCN 107
DB 5 KFVNQHLGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYQLENYCN 57
RESULT 15
AAR79056
ID AAR79056 standard; Protein; 160 AA.
XX
XX AAR79056;
AC
XX
XX 25-MAR-2003 (updated)
DT 24-JAN-1996 (first entry)
XX
XX Glycosylphosphatidylinositol-anchored human recombinant insulin.
DE
XX
XX GPI; glycosylphosphatidylinositol; insulin; hormone; solubilization;
KW Saccharomyces cerevisiae; anchor; Gas1; plasmid pBY40.

```

```

XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH Misc-difference 44..129
FT
FT
FT
XX
XX MO9522614-A1.
PN
XX
XX 24-AUG-1995.
PD
XX
XX 16-FEB-1995; 95WO-BR00010.
PF
XX
XX 17-FEB-1994; 94BR-0000600.
PR
XX
XX (ESCO-) ESCOLA PAULISTA MEDICINA.
PA (FINE-) FINEP FINANCIADORA ESTUDOS & PROJETOS.
XX
XX Cardoso De Almeida ML, Anarai De Castilho Valavicius ;
PI Gomes De Amorim Filho A;
XX
XX WPI; 1995-302720/39.
DR
XX
XX N-PSDB; AAQ99460.
DR
XX
XX Recombinant prodn. of proteins, e.g. insulin - by producing the
PT protein with a glycosylphosphatidyl:inositol anchor followed by
PT selective release
XX
XX Disclosure; Fig 3; 51pp; English.
PS
XX
XX Human recombinant insulin may be expressed in Saccharomyces
CC cerevisiae following linkage of the gene to the
CC glycosylphosphatidylinositol anchor. This anchoring technique can
CC provide for the release of the product in a highly specific and
CC selective manner. In addition, the recombinant protein will contain
CC an epitope which can be used in its final purification by
CC immunoaffinity. The protein product can be released by e.g. nitrous
CC deamination or treatment with neutral detergent.
CC (Updated on 25-MAR-2003 to correct PI field.)
XX
XX Sequence 160 AA;
SQ
Query Match 49.1%; Score 288.5; DB 16; Length 160;
Best Local Similarity 98.1%; Pred. No. 6.1e-17;
Matches 53; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 55 RFVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYQLENYCN 107
DB 43 RFVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYQLENYCN 96
Search completed: September 16, 2003, 12:38:26
Job time : 33.642 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

protein - protein search, using sw model

on: September 16, 2003, 12:37:10 : Search time 17.07 Seconds  
(without alignments)  
265.217 Million cell updates/sec

file: US-09-423-100-6

sequence: 1 MFPTPLSLRFDNAMIHAHR.....IVECCCTSI...SLVLENYCN 107

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 328717 seqs, 423108:8 residues

total number of hits satisfying chosen parameters: 328717

imum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

#### Database :

1: /cgn2\_6/protdata/1/aaa/5A\_COMB.pep.\*  
2: /cgn2\_6/protdata/1/aaa/5B\_COMB.pep.\*  
3: /cgn2\_6/protdata/1/aaa/6A\_COMB.pep.\*  
4: /cgn2\_6/protdata/1/aaa/6B\_COMB.pep.\*  
5: /cgn2\_6/protdata/1/aaa/PTCUS\_COMB.pep.\*  
6: /cgn2\_6/protdata/1/aaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	Pos	ID	Description
1	304	51.8	63	1	US-08-160-376A-6	Sequence 6, Appl
2	302.5	51.5	137	1	US-08-400-256-39	Sequence 39, Appl
3	302.5	51.5	137	3	US-08-975-365-39	Sequence 39, Appl
4	299.5	51.0	66	1	US-08-291-0608-5	Sequence 5, Appl
5	299	50.9	56	1	US-08-160-376A-7	Sequence 7, Appl
6	299	50.9	56	1	US-08-389-487-11	Sequence 11, Appl
7	299	50.9	96	1	US-08-160-376A-5	Sequence 5, Appl
8	299	50.9	96	1	US-08-389-487-8	Sequence 8, Appl
9	299	50.9	145	1	US-08-400-256-45	Sequence 45, Appl
10	299	50.9	145	3	US-08-975-365-45	Sequence 45, Appl
11	299	50.9	146	1	US-08-400-256-48	Sequence 48, Appl
12	299	50.9	146	3	US-08-975-365-48	Sequence 48, Appl
13	293	49.9	57	1	US-08-030-731A-44	Sequence 44, Appl
14	287	48.9	65	3	US-08-900-574-3	Sequence 3, Appl
15	286.5	48.8	66	3	US-08-900-574-5	Sequence 5, Appl
16	286	48.7	67	3	US-08-900-574-7	Sequence 7, Appl
17	284.5	48.5	65	1	US-08-468-674B-71	Sequence 71, Appl
18	284.5	48.5	65	1	US-08-780-571-71	Sequence 71, Appl
19	284.5	48.5	124	3	US-08-012-669F-36	Sequence 36, Appl
20	284	48.4	138	3	US-08-932-082-19	Sequence 19, Appl
21	284	48.4	140	1	US-08-400-256-33	Sequence 33, Appl
22	284	48.4	140	1	US-08-400-256-42	Sequence 42, Appl
23	284	48.4	140	3	US-08-975-365-33	Sequence 33, Appl
24	284	48.4	140	3	US-08-975-365-42	Sequence 42, Appl
25	283.5	48.3	53	1	US-08-233-617-4	Sequence 4, Appl
26	283.5	48.3	53	4	US-08-981-988A-42	Sequence 42, Appl
27	283.5	48.3	117	3	US-09-012-669F-37	Sequence 37, Appl

28	281	47.9	104	1	US-08-400-256-15	Sequence 15, Appl
29	281	47.9	104	3	US-08-975-365-15	Sequence 15, Appl
30	280.5	47.8	89	1	US-08-468-674B-41	Sequence 41, Appl
31	280.5	47.8	89	1	US-08-780-571-41	Sequence 41, Appl
32	280.5	47.8	91	1	US-08-468-674B-45	Sequence 45, Appl
33	280.5	47.8	91	1	US-08-780-571-45	Sequence 45, Appl
34	280.5	47.8	124	1	US-08-446-646-3	Sequence 3, Appl
35	279.5	47.6	167	1	US-07-518-953-8	Sequence 8, Appl
36	279.5	47.6	167	1	US-08-081-661-8	Sequence 8, Appl
37	278.5	47.4	51	4	US-09-477-924-3	Sequence 3, Appl
38	278.5	47.4	51	4	US-09-723-896-3	Sequence 3, Appl
39	278.5	47.4	51	4	US-09-280-030-63	Sequence 63, Appl
40	278	47.4	117	4	US-09-280-030-63	Sequence 63, Appl
41	277.5	47.3	53	1	US-08-233-617-3	Sequence 1, Appl
42	277	47.2	96	2	US-09-134-836-4	Sequence 4, Appl
43	277	47.2	96	4	US-09-386-303A-4	Sequence 4, Appl
44	277	47.2	97	1	US-08-160-376A-4	Sequence 4, Appl
45	277	47.2	102	1	US-08-400-256-36	Sequence 36, Appl

#### ALIGNMENTS

#### RESULT 1

US-08-160-376A-6  
; Sequence 6, Application US/08160376A  
; Patent No. 5473049  
; GENERAL INFORMATION:  
; APPLICANT: Obermeier, Kanier  
; APPLICANT: Gerl, Martin  
; APPLICANT: Ludwig, Jurgen  
; APPLICANT: Sabel, Walter  
; TITLE OF INVENTION: Process For Obtaining Proinsulin  
; TITLE OF INVENTION: Possessing Correctly Linked  
; TITLE OF INVENTION: Cystine Bridges  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenneth A. Genoni, Esq.  
; STREET: Rt. 202-206 No. 5473049H/P.O. Box 2500  
; CITY: Somerville  
; STATE: New Jersey  
; COUNTRY: U.S.A.  
; ZIP: 08876-1258  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
; COMPUTER: IBM 386  
; OPERATING SYSTEM: WINDOWS 3.1  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08160.376A  
; FILING DATE: December 1, 1993  
; CLASSIFICATION: 510  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GE P 4240420.7  
; FILING DATE: December 2, 1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Barbara V. Maurer, Esq.  
; REGISTRATION NUMBER: 31,287  
; REFERENCE/DOCKET NUMBER: HOE 92/F 384  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (908) 231-4079  
; TELEFAX: (908) 231-2255  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 63 Amino Acids  
; TYPE: Amino Acid (AA)  
; TOPOLOGY: not relevant  
US-08-160-376A-6

Query Match 51.8% Score 304; DB 1; Length 63;  
Best Local Similarity 94.7%; Pred No. 7.3e-29;  
Matches 54; Conservative 0; Mismatches 3; Indels 0; Gaps 0;



COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/291.0608  
FILING DATE: 08-AUG-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Einaudi, Carol P.  
REGISTRATION NUMBER: 32,220  
REFERENCE/DOCKET NUMBER: 02481.1105-02000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4366  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 66 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-291-0608-5

Query Match 51.0%; Score 299.5; DB 1; Length 66;  
Best Local Similarity 91.7%; Pred. No. 2.6e-28;  
Matches 55; Conservative 1; Mismatches 3; Indels 1; Gaps 1;  
QY 48 NPLGTGPRFVNHLGCSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 107  
DB 8 DPNSNG-RFVNHLGCSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 66

RESULT 5  
US-08-160-376A-7  
Sequence 7, Application US/08160376A  
Patent No. 5473049  
GENERAL INFORMATION:  
APPLICANT: Obermeier, Ranier  
APPLICANT: Gerl, Martin  
APPLICANT: Ludwig, Jorgen  
APPLICANT: Sabel, Walter  
TITLE OF INVENTION: Process For Obtaining Proinsulin  
TITLE OF INVENTION: Possessing Correctly Linked  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kenneth A. Genoni, Esq.  
STREET: Rt. 202-206 No. 5473049th/P.O. Box 2500  
CITY: Somerville  
STATE: New Jersey  
COUNTRY: U.S.A.  
ZIP: 08876-1258  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
COMPUTER: IBM 386  
OPERATING SYSTEM: WINDOWS 3.1  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/160.376A  
FILING DATE: December 1, 1993  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GE P 4240420.7  
FILING DATE: December 2, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Barbara V. Maurer, Esq.  
REGISTRATION NUMBER: 31,287  
REFERENCE/DOCKET NUMBER: HOE 92/F 384  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (908) 231-4079  
TELEFAX: (908) 231-2255  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 56 Amino Acids  
TYPE: Amino Acid (AA)  
TOPOLOGY: not relevant  
US-08-160-376A-7

Query Match 50.9%; Score 299; DB 1; Length 56;  
Best Local Similarity 100.0%; Pred. No. 2.4e-28;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 RFVNOHLGCSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 107  
DB 4 RFVNOHLGCSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 56

RESULT 6  
US-08-389-487-11  
Sequence 11, Application US/08 487  
Patent No. 5663291  
GENERAL INFORMATION:  
APPLICANT: Obermeier, Ranier  
APPLICANT: Gerl, Martin  
APPLICANT: Ludwig, Jorgen  
APPLICANT: Sabel, Walter  
TITLE OF INVENTION: Process for Obtaining Insulin Having  
TITLE OF INVENTION: Correctly Linked Cystine Bridges  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESSEE: Dunner  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States of America  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/389.487  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Einaudi, Carol P.  
REGISTRATION NUMBER: 32,220  
REFERENCE/DOCKET NUMBER: 02481.1124-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 56 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-389-487-11

Query Match 50.9%; Score 299; DB 1; Length 56;  
Best Local Similarity 100.0%; Pred. No. 2.4e-28;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 RFVNOHLGCSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 107  
DB 4 RFVNOHLGCSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 56

RESULT 7

US-08-160-376A-5  
; Sequence 5, Application US/08160376A  
; Patent No. 5473049  
; GENERAL INFORMATION:  
; APPLICANT: Obermeier, Rainer  
; APPLICANT: Gerl, Martin  
; APPLICANT: Ludwig, Jurgen  
; APPLICANT: Sabel, Walter  
; TITLE OF INVENTION: Process For Obtaining Proinsulin  
; TITLE OF INVENTION: Possessing Correctly Linked  
; TITLE OF INVENTION: Cystine Bridges  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenneth A. Genoni, Esq.  
; STREET: Rt. 202-206 No. 5473049th/P.O. Box 2500  
; CITY: Somerville  
; STATE: New Jersey  
; COUNTRY: U.S.A.  
; ZIP: 08876-1258  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
; COMPUTER: IBM 386  
; OPERATING SYSTEM: WINDOWS 3.1  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/160.376A  
; FILING DATE: December 1, 1993  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GE P 4240420.7  
; FILING DATE: December 2, 1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Barbara V. Maurer, Esq.  
; REGISTRATION NUMBER: 31,287  
; REFERENCE/DOCKET NUMBER: HOE 92/F 384  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (908) 231-4079  
; TELEFAX: (908) 231-2255  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 96 Amino Acids  
; TYPE: Amino Acid (AA)  
; TOPOLOGY: not relevant  
US-08-160-376A-5

Query Match 50.9% Score 299; DB 1; Length 96;  
Best Local Similarity 100.0%; Pred. No. 4.8e-28;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 RFVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 107  
|||||  
Db 44 RFVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 96

RESULT 8  
US-08-389-487-8  
; Sequence 8, Application US/08389487  
; Patent No. 5663291  
; GENERAL INFORMATION:  
; APPLICANT: Obermeier, Rainer  
; APPLICANT: Gerl, Martin  
; APPLICANT: Ludwig, Jurgen  
; APPLICANT: Sabel, Walter  
; TITLE OF INVENTION: Process for Obtaining Insulin Having  
; TITLE OF INVENTION: Correctly Linked Cystine Bridges  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
; ADDRESSEE: Dunner  
; STREET: 1300 I Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: United States of America

ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/389,487  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Einaudi, Carol P.  
; REGISTRATION NUMBER: 32,220  
; REFERENCE/DOCKET NUMBER: 02481.1424-00000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-408-4000  
; TELEFAX: 202-408-4400  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 96 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-389-487-8

Query Match 50.9% Score 299; DB 1; Length 96;  
Best Local Similarity 100.0%; Pred. No. 4.8e-28;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 RFVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 107  
|||||  
Db 44 RFVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 96

## RESULT 9

US-08-400-256-45  
; Sequence 45, Application US/08 00256  
; Patent No. 5750497

; GENERAL INFORMATION:  
; APPLICANT: Havelund, Svend  
; APPLICANT: Halstrom, John  
; APPLICANT: Jonassen, Ib  
; APPLICANT: Andersen, Asger Sloth  
; APPLICANT: Markussen, Jan  
; TITLE OF INVENTION: ACYLATED INSULIN  
; NUMBER OF SEQUENCES: 49  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 5750497o No. 5750497disk of No. 5750497th America, Inc.  
; STREET: 405 Lexington Avenue, 64th floor  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States of America  
; ZIP: 10174-6401

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/400,256  
; FILING DATE: 03-MAR-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Landiris, Elias J.  
; REGISTRATION NUMBER: 33,728  
; REFERENCE/DOCKET NUMBER: 3985.220-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-867-0123  
; TELEFAX: 212-878-9655  
; INFORMATION FOR SEQ ID NO: 45:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 145 amino acids

```
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-400-256-45

Query Match
Best Local Similarity 100.0%; Score 299; DB 1; Length 145;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 RFVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIKSLYLENYCN 107
Db 93 RFVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIKSLYLENYCN 145

RESULT 10
US-08-975-365-45
; Sequence 45, Application US/08975365
; Patent No. 6011007
; GENERAL INFORMATION:
; APPLICANT: Havelund, Svend
; APPLICANT: Halstrom, John
; APPLICANT: Jonassen, Ib
; APPLICANT: Andersen, Asger Sloth
; APPLICANT: Markussen, Jan
; TITLE OF INVENTION: ACYLATED INSULIN
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 57504970 No. 5750497disk of No. 6011007th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 03-MAR-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3985.220-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 145 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-400-256-48

Query Match
Best Local Similarity 100.0%; Score 299; DB 1; Length 146;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 RFVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIKSLYLENYCN 107
Db 94 RFVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIKSLYLENYCN 146

RESULT 12
US-08-975-365-48
; Sequence 48, Application US/08975365
; Patent No. 6011007
; GENERAL INFORMATION:
; APPLICANT: Havelund, Svend
; APPLICANT: Halstrom, John
; APPLICANT: Jonassen, Ib
; APPLICANT: Andersen, Asger Sloth
; APPLICANT: Markussen, Jan
; TITLE OF INVENTION: ACYLATED INSULIN
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 60110070 No. 6011007disk of No. 6011007th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: United States of America
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

APPLICATION NUMBER: US/08/975,365  
FILING DATE: 05-MAY-1987  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 38 37 273.8  
FILING DATE: 03-NOV-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 39 27 449.7  
FILING DATE: 19-AUG-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 40 12 818.0  
FILING DATE: 21-APR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Kirschner Michael K.  
REGISTRATION NUMBER: 34,851  
REFERENCE/DOCKET NUMBER: 02481-0593-02000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 44:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 57 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-030-731A-44

Query Match 50.9%; Score 299; DB 3; Length 146;  
Best Local Similarity 100.0%; Pred. No. 8e-28;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 RFVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 107  
DB 94 RFVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 146

RESULT 13  
US-08-030-731A-44  
Sequence 44, Application US/08030731A  
Patent No. 5426036  
GENERAL INFORMATION:  
APPLICANT: Koller, Klaus-Peter  
APPLICANT: Riess, Guenther Johannes  
APPLICANT: Uhlmann, Eugen  
APPLICANT: Wallmeier, Holger  
TITLE OF INVENTION: Processes for the Preparation of Foreign  
TITLE OF INVENTION: Proteins in Streptomyces  
NUMBER OF SEQUENCES: 48  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/030,731A  
FILING DATE: 12-MAR-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/189,840  
FILING DATE: 03-MAY-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/430,622  
FILING DATE: 01-NOV-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/687,610  
FILING DATE: 19-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/735,757  
FILING DATE: 29-JUL-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 37 14 866.4

FILING DATE: 05-MAY-1987  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 38 37 273.8  
FILING DATE: 03-NOV-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 39 27 449.7  
FILING DATE: 19-AUG-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 40 12 818.0  
FILING DATE: 21-APR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Kirschner Michael K.  
REGISTRATION NUMBER: 34,851  
REFERENCE/DOCKET NUMBER: 02481-0593-02000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 44:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 57 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-030-731A-44

Query Match 49.9%; Score 293; DB 1; Length 57;  
Best Local Similarity 96.2%; Pred. No. 1.3e-27;  
Matches 51; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 55 RFVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 107  
DB 5 RFVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 57

RESULT 14  
US-08-900-574-3  
Sequence 3, Application US/08900574  
Patent No. 6221837  
GENERAL INFORMATION:  
APPLICANT: Ertl, Johann  
APPLICANT: Habermann, Paul  
APPLICANT: Geisen, Karl  
APPLICANT: Seipke, Gerhard  
TITLE OF INVENTION: Insulin derivatives with increased zinc  
TITLE OF INVENTION: binding  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett,  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: District of Columbia  
COUNTRY: U.S.A.  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/900,574  
FILING DATE: July 24, 1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: German Application No. 6221837 19630242.0  
FILING DATE: July 26, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Carol P. Einaudi  
REGISTRATION NUMBER: 32,220  
REFERENCE/DOCKET NUMBER: 02481.1499-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 65 amino acids  
TYPE: Amino acid  
STRANDEDNESS: Single  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
ORGANISM: Escherichia coli  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..65  
US-08-900-574-3

Query Match 48.9%; Score 287; DB 3; Length 65;  
Best Local Similarity 91.4%; Pred. No. 7.7e-27;  
Matches 53; Conservative 0; Mismatches 3; Indels 2; Gaps 1;  
QY 51 GTGPRFVNQHLGSGHLVEALYLVCGERGFFYTPKT--RGIVEQCCTSIICSLYQLENYC 106  
Db 7 GNSARFVNQHLGSGHLVEALYLVCGERGFFYTPKTHRHGIVEQCCTSIICSLYQLENYC 64

RESULT 15  
US-08-900-574-5  
Sequence 5, Application US/08900574  
Patent No. 6221837  
GENERAL INFORMATION:  
APPLICANT: Ertl, Johann  
APPLICANT: Habermann, Paul  
APPLICANT: Geisen, Karl  
APPLICANT: Seipke, Gerhard  
TITLE OF INVENTION: Insulin derivatives with increased zinc  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett,  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: District of Columbia  
COUNTRY: U.S.A.  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/900,574  
FILING DATE: July 24, 1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: German Application No. 6221837 19630242.0  
FILING DATE: July 26, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Carol P. Elnaudi  
REGISTRATION NUMBER: 32,220  
REFERENCE/DOCKET NUMBER: 02481.1499-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 66 amino acids  
TYPE: Amino acid  
STRANDEDNESS: Single  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
ORGANISM: Escherichia coli  
FEATURE:  
NAME/KEY: Protein

LOCATION: 1..66  
US-08-900-574-5  
Query Match 48.8%; Score 286.5; DB 3; Length 66;  
Best Local Similarity 89.8%; Pred. No. 9e-27;  
Matches 53; Conservative 0; Mismatches 3; Indels 3; Gaps 1;  
QY 51 GTGPRFVNQHLGSGHLVEALYLVCGERGFFYTPKT--RGIVEQCCTSIICSLYQLENYC 106  
Db 7 GNSARFVNQHLGSGHLVEALYLVCGERGFFYTPKTAHHRGIVEQCCTSIICSLYQLENYC 65  
Search completed: September 16, 2003, 12:41:24  
Job time : 18.07 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2003, 12:38:55 : Search time 33.7237 Seconds  
(without alignments)  
472.415 Million cell updates/sec

**Title:** US-09-423-100-6

**Perfect score:** 587

**Sequence:** 1 MEPTPLSRLEFDNMLRAHR.....IVEQCCTSLCYLQLENYCN 107

**Scoring table:** BLOSUM62

Gapop 10.0 , Gapext 0.5

**Searched:** 556269 seqs, 14893369 residues

Total number of hits satisfying chosen parameters: 556269

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

**Database:** 1: **Published-Applications-AA:**

2: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	587	100.0	107	14	US-10-054-873-6
2	555.5	94.6	150	14	US-10-054-873-7
3	294	50.1	52	14	US-10-054-873-5
4	284.5	48.5	124	10	US-09-894-711-18
5	284	48.4	138	10	US-09-861-687-19
6	278.5	47.4	51	11	US-09-858-935B-5
7	278.5	47.4	51	14	US-10-028-410-3
8	278	47.4	117	9	US-09-280-030-63
9	277	47.2	96	10	US-09-947-563-4
10	275.5	46.9	124	9	US-09-736-611-12
11	275.5	46.9	124	9	US-09-740-359-12
12	275.5	46.9	124	10	US-09-894-711-12
13	275.5	46.9	125	9	US-09-736-611-10
14	275.5	46.9	125	9	US-09-740-359-10
15	275.5	46.9	125	10	US-09-894-711-10

16 275.5 46.9 147 9 US-09-736-611-8  
17 275.5 46.9 147 9 US-09-740-359-7  
18 274 46.7 144 9 US-09-736-611-6  
19 274 46.7 144 9 US-09-740-359-5  
20 274 46.7 146 10 US-09-894-711-5  
21 273 46.5 50 14 US-10-066-009A-3  
22 271 46.2 96 10 US-09-947-563-5  
23 269.5 45.9 130 9 US-09-280-030-62  
24 267 45.5 86 11 US-09-878-380-1  
25 267 45.5 86 11 US-09-858-935B-4  
26 267 45.5 86 11 US-10-028-410-2  
27 267 45.5 86 14 US-10-054-873-4  
28 267 45.5 110 9 US-09-205-658-125  
29 267 45.5 110 9 US-09-815-229-3  
30 267 45.5 110 10 US-09-804-409A-9  
31 267 45.5 110 12 US-09-969-748C-6  
32 267 45.5 110 15 US-10-038-686-1  
33 267 45.5 110 15 US-10-328-813-2  
34 260 44.3 49 14 US-10-054-873-1  
35 260 44.3 92 14 US-10-054-873-2  
36 255.5 43.5 191 11 US-09-984-010-23  
37 255.5 43.5 191 12 US-10-153-207-1  
38 255.5 43.5 191 11 US-10-400-377-1  
39 255.5 43.5 191 11 US-10-400-708-1  
40 255.5 43.5 191 11 US-10-298-148-1  
41 255.5 43.5 214 11 US-10-153-207-6  
42 255.5 43.5 217 9 US-09-929-918-9  
43 255.5 43.5 245 9 US-09-280-030-66  
44 253.5 43.2 144 9 US-09-894-711-7  
45 249.5 42.5 217 9 US-09-853-688-2

#### ALIGNMENTS

#### RESULT 1

US-10-054-873-6  
: Sequence 6, Application US/1004873  
: Publication No. US20020164712A1  
: GENERAL INFORMATION:  
: APPLICANT: Gan, Zhong Ru  
: TITLE OF INVENTION: Chimeric Protein Containing an Intramolecular Chapterone-Like Sequence

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, Eighth Floor

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/054,873

FILING DATE: 22-Jan-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/CN98/00052

FILING DATE: 31-MAR-1998

APPLICATION NUMBER: US 09/423,100

FILING DATE: 11-DEC-2000

ATTORNEY/AGENT INFORMATION:

NAME: Mycroft, Frank J

REGISTRATION NUMBER: 46,946

REFERENCE/DOCKET NUMBER: 020167-00013005

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 107 amino acids

TYPE: amino acid



```
; PRIOR APPLICATION NUMBER: PA 2000 00443
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: PA 1999 01869
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 60/211,081
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/181,450
; PRIOR FILING DATE: 2000-02-10
; PRIOR APPLICATION NUMBER: 09/740,359
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-894-711-18

Query Match      48.5%   Score 284.5; DB 10; Length 124;
Best Local Similarity 92.7%   Pred. No. 1.7e-26;
Matches 51; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

QY      54  PRFVNOHLCGSHLVEALYLVCGERGFFYTPK-TRGIVEQCCTSCISLYQLENYCN 107
Db      70  PKFVNQHLGSHLVEALYLVCGERGFFYTPKAAKGIVEQCCTSCISLYQLENYCN 124

RESULT 5
US-09-861-687-19
; Sequence 19, Application US/09861687
; Publication No. US20020193292A1
; GENERAL INFORMATION:
; APPLICANT: Markussen, Jan
; Jonassen, Ib
; Havelund, Svend
; Brandt, Jakob
; Kurtzhals, Peter
; Hansen, Hertz Per
; Kaarsholm, Niels Christian
; TITLE OF INVENTION: INSULIN DERIVATIVES
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. US20020193292A1o No. US /020193292A1disk of No. US200201932
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-Dos
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/861,687
; FILING DATE: 21-May-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/932,082
; FILING DATE: 16-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4341.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 138 amino acids
; TYPE: amino acid
```

```
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-09-861-687-19

Query Match      48.4%   Score 284; DB 10; Length 138;
Best Local Similarity 48.2%   Pred. No. 2.2e-26;
Matches 68; Conservative 5; Mismatches 28; Indels 40; Gaps 5;

QY      2  FPTIPLSRFLFDNMLRAHRLHQLAFDTYOFFEAYTPKEQ--KVSFLO-----N 48
Db      3  FPSI-----FTAVLFAASSALAAPVNTTDEIAQIPAEAVIGYSNLEGGDFVAVLPFSN 57
QY      49  PLGTG-----PRFVNOHLCGSHLVEALYLVCGERGFFYTPK-IR 86
Db      58  STNNGLLFINTTIASTAARKEGVSLDKRFVNQHLGSHLVEALYLVCGERGFFYTPKAAK 117
QY      87  GIVEQCCTSCISLYQLENYCN 107
Db      118  GIVEQCCTSCISLYQLENYCN 138

RESULT 6
US-09-858-935B-5
; Sequence 5, Application US/09858935B
; Publication No. US200300691777
; GENERAL INFORMATION:
; APPLICANT: Dubaquié, Yves
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Lowman, Henry B.
; TITLE OF INVENTION: METHOD FOR TREATING CARTILAGE DISORDERS
; FILE REFERENCE: P1794R1
; CURRENT APPLICATION NUMBER: US/09/858,935B
; CURRENT FILING DATE: 2002-07-02
; PRIOR APPLICATION NUMBER: US 60/248,985
; PRIOR FILING DATE: 2000-11-15
; PRIOR APPLICATION NUMBER: US 60/204,490
; PRIOR FILING DATE: 2000-05-16
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 1
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-858-935B-5

Query Match      47.4%   Score 278.5; DB 11; Length 51;
Best Local Similarity 98.1%   Pred. No. 2.9e-26;
Matches 51; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY      56  FVNQHLGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYQLENYCN 107
Db      1  FVNQHLGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYQLENYCN 51

RESULT 7
US-10-028-410-3
; Sequence 3, Application US/10028410
; Publication No. US20020160955A1
; GENERAL INFORMATION:
; APPLICANT: Dubaquié, Yves
; APPLICANT: Lowman, Henry
; TITLE OF INVENTION: PROTEIN VARIANTS
; FILE REFERENCE: P1712R1-1
; CURRENT APPLICATION NUMBER: US/10/028,410
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: US/09/477,924
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 3
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-028-410-3
```

```
Query Match          47.4%; Score 278.5; DB 14; Length 51;
Best Local Similarity 98.1%; Pred. No. 2.9e-26;
Matches 51; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 56 FVNHLCGSHLVEALYVCGERGFFYTPKTRGIVEOCCT-ICSLYOLENYCN 107
      |||||
Db 1 FVNHLCGSHLVEALYVCGERGFFYTPKT-GIVEOCCT-ICSLYOLENYCN 51

RESULT 8
US-09-280-030-63
; Sequence 63, Application US/09280030A
; Patent No. US20010021515A1
; GENERAL INFORMATION:
; APPLICANT: Sato, Seiji
; APPLICANT: Higashikuni, Naohiko
; APPLICANT: Kudo, Toshiyuki
; APPLICANT: Kondo, Masaaki
; TITLE OF INVENTION: DNAS ENCODING NEW FUSION PROTEINS AND PROCESSES FOR
; TITLE OF INVENTION: PREPARING USEFUL POLYPEPTIDES THROUGH EXPRESSION OF THE
; FILE REFERENCE: 382.1026
; CURRENT APPLICATION NUMBER: US/09/280.030A
; EARLIER FILING DATE: 1999-03-26
; EARLIER FILING DATE: JP10-87339/1998
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 63
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Designated as
; OTHER INFORMATION: an amino acid sequence of
; OTHER INFORMATION: MWpmp10-Met-Proinsulin
US-09-280-030-63

Query Match          47.4%; Score 278; DB 9; Length 117;
Best Local Similarity 60.0%; Pred. No. 9.5e-26;
Matches 54; Conservative 0; Mismatches 2; Indels 34; Gaps 1;

Qy 52 TGRFVNHLCGSHLVEALYVCGERGFFYTPKT----- 85
      |||||
Db 28 TAPFVNHLCGSHLVEALYVCGERGFFYTPKTRREADMLVQGVELGGPGAGSLQPL 87

Qy 86 -----RGIVEQCCT-ICSLYOLENYCN 107
      |||||
Db 88 ALGSLQKRGIVEQCCT-ICSLYOLENYCN 117

RESULT 9
US-09-947-563-4
; Sequence 4, Application US/09947563
; Patent No. US20020156234A1
; GENERAL INFORMATION:
; APPLICANT: Rubroder, Franz-Josef
; APPLICANT: Keller, Reinhold
; TITLE OF INVENTION: Improved process for obtaining
; insulins precursors having correctly bonded cystine bridges
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farrahaw, Garrett &
; Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/947,563
FILING DATE: 07-Sep-2001
CLASSIFICATION: <Unknown>
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 09/134,836
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Leslie McDonnell
REGISTRATION NUMBER: 34,872
REFERENCE/DOCKET NUMBER: 02481.1600-000000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4900
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 96 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: protein
LOCATION: 1..96
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-947-563-4

Query Match          47.2%; Score 277; DB 10; Length 96;
Best Local Similarity 59.3%; Pred. No. 9.8e-26;
Matches 54; Conservative 0; Mismatches 3; Indels 34; Gaps 1;

Qy 51 GTGPFVNHLCGSHLVEALYVCGERGFFYTPKT----- 85
      |||||
Db 6 GNSRFVNHLCGSHLVEALYVCGERGFFYTPKTRREADMLVQGVELGGPGAGSLQPL 65

Qy 86 -----RGIVEQCCT-ICSLYOLENYCN 107
      |||||
Db 66 IALEGSLQKRGIVEQCCT-ICSLYOLENYCN 96

RESULT 10
US-09-736-611-12
; Sequence 12, Application US/09736611
; Patent No. US20010023069A1
; GENERAL INFORMATION:
; APPLICANT: Kjeldsen, Thomas
; APPLICANT: Ludvigsen, Svend
; APPLICANT: Kaarsholm, Niels
; TITLE OF INVENTION: Method For Making Insulin Precursors and
; FILE REFERENCE: 6058.200-US
; CURRENT APPLICATION NUMBER: US/09/736,611
; PRIOR FILING DATE: 2000-12-14
; PRIOR APPLICATION NUMBER: 60/181,443
; PRIOR FILING DATE: 2000-02-10
; PRIOR APPLICATION NUMBER: 60/211,441
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: PA 1999 01868
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: PA 000 00440
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 124
; TYPE: PRT
; ORGANISM: N-terminal extension
US-09-736-611-12
```

```

Query Match      46.9%:  Score 275.5;  DB 9;  Length 124;
Best Local Similarity  90.9%:  Pred. No. 2.le-25;
Matches 50;  Conservative 2;  Mismatches 2;  Indels 1;  Gaps 1;

OY  54  PAFVNHQHGSHLVREALYLVGGERGFFYTPKT-RCGIVEQCCTSCISLYOLENYCN 107
      1-|||||1|||||1|||||1|||||1|||||1|||||1|||||1|||||1|||||1|||||1
      70  PAFVNHQHGSHLVREALYLVGGERGFFYTDKRGKIVEQCCTSCISLYOLENYCN 124

RESULT 11
US-09-740-359-12
; Sequence 12, Application US/09740359
; Patent No. US20010041787A1
; GENERAL INFORMATION:
; APPLICANT: Kjeldsen, Thomas Borglum
; APPLICANT: Ludvigsen, Svend
; TITLE OF INVENTION: Method for making insulin precursors and
; TITLE OF INVENTION: Insulin precursor analogues having improved fermentation
; TITLE OF INVENTION: Insulin precursor analogues having improved fermentation
; FILE REFERENCE: 6148.200-US
; CURRENT APPLICATION NUMBER: US/09/740,359
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: PA 2000 00443
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: PA 1999 01869
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 60/211,081
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/181,450
; PRIOR FILING DATE: 2000-02-10
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 124
; TYPE: PRT
; ORGANISM: T457 leader fused with N-terminally extended....
US-09-740-359-12

Query Match      46.9%:  Score 275.5;  DB 9;  Length 124;
Best Local Similarity  90.9%:  Pred. No. 2.le-25;
Matches 50;  Conservative 2;  Mismatches 2;  Indels 1;  Gaps 1;

OY  54  PAFVNHQHGSHLVREALYLVGGERGFFYTPKT-RCGIVEQCCTSCISLYOLENYCN 107
      1-|||||1|||||1|||||1|||||1|||||1|||||1|||||1|||||1|||||1|||||1
      70  PAFVNHQHGSHLVREALYLVGGERGFFYTDKRGKIVEQCCTSCISLYOLENYCN 124

RESULT 12
US-09-894-711-12
; Sequence 12, Application US/09894711
; Patent No. US20020137144A1
; GENERAL INFORMATION:
; APPLICANT: Kjeldsen, Thomas Borglum
; APPLICANT: Ludvigsen, Svend
; TITLE OF INVENTION: Method for making insulin precursors and
; TITLE OF INVENTION: Insulin precursor analogues having improved fermentation
; TITLE OF INVENTION: Insulin precursor analogues having improved fermentation
; FILE REFERENCE: 6148.400-US
; CURRENT APPLICATION NUMBER: US/09/894,711
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: PA 2000 00443
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: PA 1999 01869
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 60/211,081
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/181,450
; PRIOR FILING DATE: 2000-02-10
; PRIOR APPLICATION NUMBER: 09/740,359
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12

```

; PRIOR APPLICATION NUMBER: 60/181,450
; PRIOR FILING DATE: 2000-02-10
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 125
; TYPE: PRT
; ORGANISM: TA39 leader fused with N-terminally extended
US-09-740-359-10

Query Match 46.9% Score 275.5; DB 9; Length 125;
Best Local Similarity 90.9% Pred. No. 2.1e-25;
Matches 50; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 54 PRFVNQHLGSHLVEALYLVCGERGFFYTPKT-RGIVEQCCTSIICSLYQLENYCN 107
I:|||||
DB 71 PRFVNQHLGSHLVEALYLVCGERGFFYTDKDGKGIIVEQCCTSIICSLYQLENYCN 125
I:|||||

RESULT 15
US-09-894-711-10
; Sequence 10, Application US/09894711
; Patent No. US20020137144A1
; GENERAL INFORMATION:
; APPLICANT: Kjeldsen, Thomas Borglum
; APPLICANT: Ludvigsen, Svend
; TITLE OF INVENTION: Method for making insulin precursors and
; TITLE OF INVENTION: insulin precursor analogues having improved fermentation
; TITLE OF INVENTION: yield in yeast
; FILE REFERENCE: 6148.400-US
; CURRENT APPLICATION NUMBER: US/09/894,711
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: PA 2000 00443
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: PA 1999 01869
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 60/211,081
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/181,450
; PRIOR FILING DATE: 2000-02-10
; PRIOR APPLICATION NUMBER: 09/740,359
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 125
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-894-711-10

Query Match 46.9% Score 275.5; DB 10; Length 125;
Best Local Similarity 90.9% Pred. No. 2.1e-25;
Matches 50; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 54 PRFVNQHLGSHLVEALYLVCGERGFFYTPKT-RGIVEQCCTSIICSLYQLENYCN 107
I:|||||
DB 71 PRFVNQHLGSHLVEALYLVCGERGFFYTDKDGKGIIVEQCCTSIICSLYQLENYCN 125
I:|||||

Search completed: September 16, 2003, 12:52:26
Job time : 34.7237 secs

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	275.5	46.8	96	2	PC7082	epidermal growth f
2	273.5	46.6	51	1	INMHP	insulin - sperm wh
3	273.5	46.6	51	1	INMHP	insulin - finback
4	273.5	46.6	51	1	INFL	insulin - elephant
5	273	46.5	110	2	B42179	insulin precursor
6	273	46.5	110	2	QJ0178	insulin precursor
7	271.5	46.3	51	1	INMY	insulin - hamster
8	268.5	45.7	51	1	INMSSP	insulin - Egyptian
9	267.5	45.6	51	2	A59151	insulin precursor
10	267	45.5	110	1	IPHU	insulin precursor
11	267	45.5	110	2	A42179	insulin precursor
12	263.5	44.9	51	1	INMH15	insulin - sei whale
13	263.5	44.9	51	1	INGF	insulin - goat
14	263.5	44.9	51	1	INCM4	insulin - Arabian
15	263	44.8	84	1	IPGC	insulin precursor
16	263	44.8	110	1	INBB	insulin precursor
17	262.5	44.7	51	1	INCT	insulin - cat
18	262	44.6	110	1	IPGQ	insulin precursor
19	261.5	44.5	51	1	INMKSQ	insulin - common s
20	260	44.3	110	2	I48166	insulin precursor
21	258.5	44.0	105	1	IPBO	insulin precursor
22	257	43.8	108	2	A39883	insulin precursor
23	256.5	43.7	51	2	QJ0362	insulin - North Am
24	255.5	43.5	217	1	STHU	somatotropin - pre
25	255.5	43.5	217	2	I67410	somatotropin - rhe
26	252.5	43.0	77	1	INSH	insulin precursor
27	252	42.9	86	1	IPHO	insulin precursor
28	251.5	42.8	51	1	INCB	insulin - Chinchil
29	250	42.6	108	1	INMS1	insulin l precursor



Db 23 PAFVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAEDQVGVQLGGPGAGSLQPLAL 82

QY 86 -----RGIVEQCCTSIICSLYLENYCN 107

Db 83 EGSLOKRGIVEQCCTSIICSLYLENYCN 110

# RESULT 7

INHY insulin - hamster

C:Species: Cricetinae gen. sp. (hamster)

C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 16-Jul-1999

C:Accession: A91456

R:Neelon, F.A.; Delcher, H.K.; Steinman, H.; Lebovitz, H.E.

Ref. Proc. 32, 300, 1973

A:Title: Structure of hamster insulin: comparison with a tumor insulin.

A:Reference number: A91456

A:Accession: A91456

A:Molecule type: protein

A:Residues: 1-30; 31-51 <NEE>

C:Superfamily: Insulin

C:Keywords: hormone; pancreas

F:1-30/Domain: insulin chain B #status experimental <BCH>

F:1-30;31-51/Product: insulin #status experimental <MAT>

F:31-51/Domain: insulin chain A #status experimental <ACH>

F:7-37;19-50;36-41/Disulfide bonds: #status predicted

Query Match 46.3%; Score 271.5; DB 1; Length 51;  
Best Local Similarity 94.2%; Pred. No. 1.4e-21;  
Matches 49; Conservative 2; Mismatches 0; Indels 1; Gaps 1;

QY 56 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIICSLYLENYCN 107

Db 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKS-GIVDQCCTSIICSLYLENYCN 51

# RESULT 8

INWSSP insulin - Egyptian splny mouse (tentative sequence)

C:Species: Acomys cahirinus (Egyptian splny mouse)

C:Date: 13-Jul-1981 #sequence\_revision 13-Jul-1981 #text\_change 31-Mar-2000

C:Accession: A01591

R:Buenzli, H.F.; Humbel, R.E.

Hoppe-Seyler's Z. Physiol. Chem. 353, 444-450, 1972

A:Title: Isolation and partial structural analysis of insulin from mouse (Mus musculus)

A:Reference number: A01591; MUID:72189454; PMID:5028210

A:Contents: composition

A:Accession: A01591

A:Molecule type: protein

A:Residues: 1-30; 31-51 <BUE>

C:Superfamily: Insulin

C:Keywords: hormone; pancreas

F:1-30/Domain: insulin chain B #status predicted <BCH>

F:1-30;31-51/Product: insulin #status predicted <MAT>

F:31-51/Domain: insulin chain A #status predicted <ACH>

F:7-37;19-50;36-41/Disulfide bonds: #status predicted

Query Match 45.7%; Score 268.5; DB 1; Length 51;  
Best Local Similarity 92.3%; Pred. No. 2.9e-21;  
Matches 48; Conservative 3; Mismatches 0; Indels 1; Gaps 1;

QY 56 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIICSLYLENYCN 107

Db 1 FVBOHLCGSHLVEALYLVCGERGFFYTPKS-GIVDQCCTSIICSLYLENYCN 51

# RESULT 9

A59151 insulin precursor - jack bean (fragments)

N:Alternate names: hypoglycemic agent; plant insulin

C:Species: Canavalia ensiformis (jack bean)

C:Date: 07-Dec-1999 #sequence\_revision 07-Dec-1999 #text\_change 10-Dec-1999

C:Accession: B59151; A59151

R:Oliveira, A.E.A.; Machado, O.L.T.; Gomes, V.M.; Xavier-Neto, J.; Pereira, A.C.P.; V Protein Pept. Lett. 6, 15-21, 1999

A:Title: Jack bean seed coat contains a protein with complete sequence homology to bc

A:Reference number: A59151

A:Accession: B59151

A:Molecule type: protein

A:Residues: 1-30 <MACB>

A:Accession: A59151

A:Molecule type: protein

A:Residues: 31-51 <MACA>

C:Comment: The two chains are probably produced from the same precursor.

C:Superfamily: Insulin

F:1-30;31-51/Product: insulin #status experimental <MAT>

F:1-30/Domain: chain B #status experimental <CHB>

F:31-51/Domain: chain A #status experimental <CHA>

F:7-37;19-50;36-41/Disulfide bonds: #status predicted

Query Match 45.6%; Score 267.5; DB 2; Length 51;  
Best Local Similarity 92.3%; Pred. No. 3.7e-21;  
Matches 48; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 56 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIICSLYLENYCN 107

Db 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKA-GIVEQCCTSIICSLYLENYCN 51

# RESULT 10

IPHU insulin precursor [validated] - human

N:Alternate names: preproinsulin

C:Species: Homo sapiens (man)

C:Date: 23-Oct-1981 #sequence\_revision 23-Oct-1981 #text\_change 08-Dec 2000

C:Accession: A93222; A94253; A93216; A94251; A93144; A92075; A91166; A91114; A01579;

R:Bell, G.I.; Picot, R.L.; Rutter, W.J.; Cordell, B.; Tischer, E.; Goodman, H.M.

Nature 284, 26-32, 1980

A:Title: Sequence of the human insulin gene.

A:Reference number: A93222; MUID:80120725; PMID:6243748

A:Accession: A93222

A:Molecule type: DNA

A:Residues: 1-110 <REF>

A:Cross-references: GB:J00265, D:q186429; PIDN:AAA59172.1; PID:q386828

R:Ullrich, A.; Dull, T.J.; Gray, A.; Brosius, J.; Sures, I.

Science 209, 612-615, 1980

A:Title: Genetic variation in the human insulin gene.

A:Reference number: A94253; MUID:8026313; PMID:6248962

A:Accession: A94253

A:Molecule type: DNA

A:Residues: 1-110 <ULL>

A:Cross-references: GB:J00265; MUID:q186429; PIDN:AAA59172.1; PID:q386828

R:Bell, G.I.; Swain, W.F.; Picot, R.; Cordell, B.; Goodman, H.M.; Rutter, W.J.

Nature 282, 525-527, 1979

A:Title: Nucleotide sequence of a cDNA clone encoding human preproinsulin.

A:Reference number: A93216; MUID:80054779; PMID:503234

A:Accession: A93216

A:Molecule type: mRNA

A:Residues: 1-110 <BEL>

A:Cross-references: GB:J00265; MUID:q186429; PIDN:AAA59172.1; PID:q386828

R:Sures, I.; Goeddel, D.V.; Gray, A.; Ullrich, A.

Science 208, 57-59, 1980

A:Title: Nucleotide sequence of human preproinsulin complementary DNA.

A:Reference number: A94251; MUID:80147417; PMID:6927840

A:Accession: A94251

A:Molecule type: mRNA

A:Residues: 1-110 <SUR>

A:Cross-references: GB:J00265; MUID:q186429; PIDN:AAA59172.1; PID:q386828

R:Nicol, D.S.H.W.; Smith, L.F.

Nature 187, 483-485, 1960

A:Title: Amino-acid sequence of human insulin.

A:Reference number: A93144

A:Accession: A93144

A:Molecule type: protein

A:Residues: 25-54; 90-110 <NIC>

R:Oyer, P.E.; Cho, S.; Peterson, J.D.; Steiner, D.F.

J. Biol. Chem. 246, 1375-1386, 1971  
A:Title: Studies on human proinsulin. Isolation and amino acid sequence of the human pan  
A:Reference number: A92075; MUID:71116410; PMID:5101771  
A:Accession: A92075  
A:Molecule type: protein  
A:Residues: 57-87 <OYE>  
R:Ko, A.; Smyth, D.G.; Markussen, J.; Sundby, F.  
Eur. J. Biochem. 20, 190-199, 1971  
A:Title: Amino acid sequence of the C-peptide of human proinsulin.  
A:Reference number: A91186; MUID:71257722; PMID:5560404  
A:Accession: A91186  
A:Molecule type: protein  
A:Residues: 57-87 <KOA>  
R:Lucassen, A.M.; Julier, C.; Beressi, J.P.; Boitard, C.; Froquel, P.; Lathrop, M.; Bell  
Nature Genet. 4, 305-310, 1993  
A:Title: Susceptibility to insulin dependent diabetes mellitus maps to a 4.1 kb segment  
A:Reference number: I58114; MUID:93364428; PMID:8358440  
A:Accession: I58114  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-59,63-110 <RES>  
A:Cross-references: GB:L15440; NID:q307071; PIDN:AAA59179.1; PID:q307072  
R:Sieber, P.; Kamber, B.; Hartmann, A.; Joehl, A.; Riniker, B.; Rittel, W.  
Helv. Chim. Acta 57, 2617-2621, 1974  
A:Title: Totalsynthese von Humaninsulin unter gezielter Bildung der Disulfidbindungen.  
A:Reference number: A91636; MUID:75077277; PMID:4443293  
A:Contents: annotation; synthesis  
A:Note: disulfide-bonded human insulin was synthesized; the synthetic hormone was identical  
A:Note: article in German with English abstract  
R:Naithani, V.K.  
Hoppe-Seyler's Z. Physiol. Chem. 354, 659-672, 1973  
A:Title: The synthesis of C-peptide of human proinsulin.  
A:Reference number: A91638; MUID:75040007; PMID:4803504  
A:Contents: annotation; synthesis of residues 57-87  
R:Geiger, R.; Jaeger, G.; Koenig, W.  
Chem. Ber. 106, 2347-2352, 1973  
A:Title: Synthesis of the complete sequence of human proinsulin C-peptide and its [Glu-9  
A:Reference number: A90914  
A:Contents: annotation; synthesis of residues 57-87  
R:Kaufmann, J.E.; Irminger, J.C.; Halban, P.A.  
Biochem. J. 310, 869-874, 1995  
A:Title: Sequence requirements for proinsulin processing at the B-chain/C-peptide juncti  
A:Reference number: S58661; MUID:96013185; PMID:7575420  
A:Contents: annotation; site-directed mutagenesis study of proteolytic processing  
C:Genetics:  
A:Gene: GDB:INS  
A:Cross-references: GDB:119349; OMIM:176730  
A:Map position: 11p15.5-11p15.5  
A:Introns: 63/1  
C:Superfamily: insulin  
C:Keywords: hormone; pancreas  
F:1-24/Domain: signal sequence \*status predicted <SIC>  
F:25-54/Domain: insulin chain B \*status experimental <BCH>  
F:55-90/Domain: insulin \*status experimental <MAT>  
F:91-110/Domain: connecting C peptide \*status experimental <CPEP>  
F:90-110/Domain: insulin chain A \*status experimental <ACH>  
F:31-96,43-109,95-100/Disulfide bonds: \*status experimental

Query Match 45.5% Score 267; DB 1; Length 110;  
Best Local Similarity 60.5% Pred. No. 9.3e-21;  
Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;  
Oy 56 FVNOHLCGSHLVEALYLVCGERGFFYTPKT----- 85  
Db 25 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRAEADLQGVQLGGPGAGSLQPLALEG 84  
Oy 86 ----RGIVEQCCTSCSLYLENYCN 107  
Db 85 SLOKRGIVEQCCTSCSLYLENYCN 110

RESULT 11  
A42179.

insulin precursor - chimpanzee  
C:Species: Pan troglodytes (chimpanzee)  
C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 16-Jul-1999  
C:Accession: A42179; S22058  
R:Seino, S.; Bell, G.I.; Li, W.H.  
Mol. Biol. Evol. 9, 193-203, 1992  
A:Title: Sequences of primate insulin genes support the hypothesis of a slower rate o  
A:Reference number: A42179; MUID:92219953; PMID:1560757  
A:Accession: A42179  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-110 <SEI>  
A:Cross-references: EMBL:X61089; NID:q38251; PIDN:CAA43403.1; PID:q38252  
A:Note: sequence extracted from NCBI backbone (NCBIP:95067)  
C:Genetics:  
A:Introns: 63/1  
C:Superfamily: insulin

Query Match 45.5% Score 267; DB 2; Length 110;  
Best Local Similarity 60.5% Pred. No. 9.3e-21;  
Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;  
Oy 56 FVNOHLCGSHLVEALYLVCGERGFFYTPKT----- 85  
Db 25 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRAEADLQGVQLGGPGAGSLQPLALEG 84  
Oy 86 ----RGIVEQCCTSCSLYLENYCN 107  
Db 85 SLOKRGIVEQCCTSCSLYLENYCN 110

RESULT 12  
INSULIN - sei whale  
C:Species: Balaeoptera borealis (sei whale)  
C:Date: 13-Jul-1981 #sequence\_revision 13-Jul-1981 #text\_change 16-Jul-1999  
C:Accession: A01582  
R:Ishihara, Y.; Saito, T.; Ito, Y.; Fujino, M.  
Nature 181, 1468-1469, 1958  
A:Title: Structure of sperm- and sei-whale insulins and their breakdown by whale peps  
A:Reference number: A93142  
A:Accession: A01582  
A:Molecule type: protein  
A:Residues: 1-30,31-51 <ISH>  
C:Superfamily: insulin  
C:Keywords: hormone; pancreas  
F:1-30/Domain: insulin chain B \*status experimental <BCH>  
F:31-30,31-51/Product: insulin \*status experimental <MAT>  
F:31-51/Domain: insulin chain A \*status experimental <ACH>  
F:7-37,19-50,36-41/Disulfide bonds: \*status predicted

Query Match 44.9% Score 263.5; DB 1; Length 51;  
Best Local Similarity 92.3% Pred. No. 9.6e-21;  
Matches 48; Conservative 0; Mismatches 3; Indels 1; Gaps 1;  
Oy 56 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCSLYLENYCN 107  
Db 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKA-GIVEQCCTSCSLYLENYCN 51

RESULT 13  
INGT  
insulin - goat  
C:Species: Capra aegagrus hircus (domestic goat)  
C:Date: 13-Jul-1981 #sequence\_revision 13-Jul-1981 #text\_change 16-Jul-1999  
C:Accession: A01586  
R:Smith, L.F.  
Am. J. Med. 40, 662-666, 1966  
A:Title: Species variation in the amino acid sequence of insulin.  
A:Reference number: A90029; MUID:66160119; PMID:5949593  
A:Accession: A01586  
A:Molecule type: protein  
A:Residues: 1-30,31-51 <SMI>

RESULT 15  
IPPC  
insulin precursor - pig  
C:Species: Sus scrofa domestica (domestic pig)  
C>Date: 22-Jun-1981 #sequence\_revision 22-Jun-1981 #text\_change 16-Jul-1999  
C:Accession: A01583; S16492; A60835; B60835  
R:Chance, R.E.; Ellis, R.M.; Bromer, W.W.  
Science 161, 165-167, 1968  
A:Title: Porcine proinsulin: characterization and amino acid sequence.  
A:Reference number: A94240; MUID:68286485; PMID:5657063  
A:Accession: A01583  
A:Molecule type: protein  
A:Residues: 1-34, 'Q', 36-84 <CHA>  
R:Chance, R.E.  
submitted to the Atlas, July 1970  
A:Reference number: A94572  
A:Accession: A94572  
A:Molecule type: protein  
A:Residues: 1-84 <CH2>  
R:Brown, H.; Sanger, F.; Kitai, R.  
Biochem. J. 60, 556-565, 1955  
A:Title: The structure of pig and sheep insulins.  
A:Reference number: A90344  
A:Accession: S16492  
A:Molecule type: protein  
A:Residues: 1-30;31-51 <BRO>  
R:Shel, L.; Damgaard, O.  
Horm. Metab. Res. 20, 476-480, 1988

GenCore version .1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2003, 12:33:30 ; Search time 8.32685 Seconds  
(without alignments)  
604.293 Million cell updates/sec

**Title:** us-09-423-100-6

**Perfect score:** 587

**Sequence:** 1 MFPTPLSRFLFNAMLAHRR.....IVEQCCTSCSYQLENYCN 107

**Scoring table:** BLOSUM62

Gapop 10.0 , Gapext 0.5

**Searched:** 127863 seqs, 47026705 residues

**Total number of hits satisfying chosen parameters:** 127863

**Minimum DB seq length:** 0

**Maximum DB seq length:** 2000000000

**Post-processing:** Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

**Database :** SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	273.5	46.6	51	INS_BALPH	P01312 balaenopter
2	273.5	46.6	51	INS_ELEMA	P01316 elephas max
3	273	46.5	110	INS_CERAE	P30407 cercoptithec
4	273	46.5	110	INS_MACFA	P30406 macaca fasc
5	268.5	45.7	51	INS_ACOCA	P01324 acomys cahi
6	267	45.5	110	INS_HUMAN	P01308 homo sapien
7	267	45.5	110	INS_PANTR	P30410 pan troglod
8	266	45.3	110	INS_SPETR	O91x13 spermophilu
9	263.5	44.9	51	INS_BALBO	P01314 balaenopter
10	263.5	44.9	51	INS_CAMDR	P01320 camelus dro
11	263.5	44.9	51	INS_CAPHI	P01319 capra hircu
12	263	44.8	108	INS_PIG	P01315 sus scrofa
13	263	44.8	110	INS_RABIT	P01311 oryctolagus
14	262.5	44.7	51	INS_FELCA	P06306 felis silve
15	262	44.6	110	INS_CANFA	P01321 canis fami
16	260	44.3	110	INS_CRILQ	P01313 cricetus
17	258.5	44.0	105	INS_BOVIN	P01317 bos taurus
18	257	43.8	108	INS_BOVIN	P10604 aotus trivi
19	257	43.8	110	INS_PSAOB	Q62587 psammomys o
20	256.5	43.7	51	INS_DIDMA	P18109 didelphis m
21	255.5	43.5	217	INS_HUMAN	P01241 homo sapien
22	255.5	43.5	217	SOMA_MACMU	P33093 macaca mula
23	255.5	43.5	217	SOMA_PANTR	P58756 pan troglod
24	254.5	43.4	105	INS_SHEEP	P01318 ovis aries
25	252	42.9	86	INS_HORSE	P01310 equus cabal
26	251.5	42.8	51	INS_CHIBR	P01327 chinchilla
27	250	42.6	108	INS_MOUSE	P01325 mus musculu
28	249	42.4	110	INSI_RAT	P01322 rattus norv
29	249	42.4	217	SOMA_CALJA	O9gmb3 callithrix
30	249	42.4	51	SOMA_SAIBB	P58343 saimir bol
31	248.5	42.3	51	INS_ANSAN	P07454 anser anser
32	248	42.2	110	INS2_MOUSE	P01326 mus musculu
33	248	42.2	110	INS2_RAT	P01323 rattus norv

34 246 41.9 52 1 INS\_ACIGU  
35 245 41.7 103 1 INS\_SELRF  
36 244.5 41.7 51 1 INS\_HYSCR  
37 244.5 41.7 51 1 INS\_TRASC  
38 239.5 40.8 107 1 INS\_CHICK  
39 236 40.2 217 1 SOM2\_PANTR  
40 235.5 40.1 51 1 INS\_ORNAN  
41 233.5 39.8 81 1 INS\_ANAPL  
42 231.5 39.4 51 1 INS\_ALIMI  
43 231 39.4 52 1 INS\_LEPSP  
44 228.5 38.9 51 1 INS\_ZAODH  
45 228 38.8 217 1 SOM2\_HUMAN

#### ALIGNMENTS

RESULT 1  
INS\_BALPH  
ID INS\_BALPH STANDARD; PRT: 51 AA.  
AC P01312:  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Insulin.  
GN INS.  
OS Balaenoptera physalus (Finback whale) (Common rorqual), and  
OS Physeter catodon (Sperm whale) (Physeter macrocephalus).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;  
OC Balaenopteridae; Balaenoptera.  
OX NCBI\_taxid=9770, 9755;  
RN [1]  
RP PARTIAL SEQUENCE.  
RC SPECIES-B.physalus;  
RA Hama H., Titani K., Sakaki S., Narita K.;  
RT "The amino acid sequence in fin-whale insulin.";  
RL J. Biochem. 56:285-293(1964)  
RN [2]  
RP SEQUENCE.  
RC SPECIES-P.catodon;  
RA Ishihara Y., Saito T., Ito Y., Fujino M.;  
RT "Structure of sperm- and sei-whale insulins and their breakdown by  
RT whale pepsin.";  
RL Nature 181:1468-1469(1958).  
RN [3]  
RP SEQUENCE.  
RC SPECIES-P.catodon;  
RA Harris J.I., Sanger F., Naughton M.A.;  
RT "Species differences in insulin.";  
RL Arch. Biochem. Biophys. 65:427-438(1956).  
CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
CC DISULFIDE BONDS.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
DR PIR: A91918; INWHF.  
DR PIR: A93142; INWHP.  
DR HSSP: P01317; LAPH.  
DR InterP: IPR004825; Ins/IGF/relax.  
DR SMART: SM00078; IIGF; 1.  
DR PROSITE: PS00262; INSULIN; 1.  
KW Insulin family; Hormone; Glucose metabolism.  
FT CHAIN 1 30 INSULIN B CHAIN.  
FT NON\_CONS 30 31  
FT CHAIN 31 51 INSULIN A CHAIN.  
FT DISULFID 7 37 INTERCHAIN.  
FT DISULFID 19 50 INTERCHAIN.  
FT DISULFID 36 41  
SQ SEQUENCE 51 AA: 5766 MW: 9007B514691A7CDD CRC64:

Query Match 46.6%; Score 273.5; DB 1; Length 51;  
 Best Local Similarity 96.2%; Pred. No. 1.6e-22;  
 Matches 50; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 56 FVNQHLGCSHLVEALYLVCGERGFFYTPKTRGIVEQCCTICSILYOLENYCN 107  
 |||||||  
 DB 1 FVNQHLGCSHLVEALYLVCGERGFFYTPKA-GIVEQCCTICSILYOLENYCN 51

## RESULT 2

INS\_ELEMA STANDARD; PRT: 51 AA.  
 AC P01316;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE Insulin.  
 GN INS.  
 OS Elephas maximus (Indian elephant).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Proboscidea; Elephantidae; Elephas.  
 OX NCBI\_TaxID=9783;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=66160119; PubMed=5949593;  
 RA Smith L.F.;  
 RL Am. J. Med. 40:662-666(1966).  
 CC \*Species variation in the amino acid sequence of insulin.\*;  
 CC \*FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
 CC \*SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
 CC DISULFIDE BONDS.  
 CC \*SUBCELLULAR LOCATION: Secreted.  
 CC \*MISCELLANEOUS: THE SPECIES OF ELEPHANT IS NOT GIVEN, BUT IT IS  
 CC MOST PROBABLY THE INDIAN ELEPHANT (ELEPHAS MAXIMUS).  
 CC \*SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 DR HSSP: P01308; IAI0.  
 DR InterPro: IPR0004825; Ins/IGF/relax.  
 DR SMART: SM00078; IIGF; 1.  
 DR PROSITE: PS00262; INSULIN; 1.  
 KW Insulin family; Hormone; Glucose metabolism.  
 FT CHAIN 1 30 INSULIN B CHAIN.  
 FT NON\_CONS 30 31  
 FT CHAIN 31 51 INSULIN A CHAIN.  
 FT DISULFID 7 37 INTERCHAIN.  
 FT DISULFID 19 50 INTERCHAIN.  
 FT DISULFID 36 41  
 SQ SEQUENCE 51 AA: 5752 MW: 9007850CDB457D6D CR014;

Query Match 46.6%; Score 273.5; DB 1; Length 51;  
 Best Local Similarity 94.2%; Pred. No. 1.6e-22;  
 Matches 49; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

QY 56 FVNQHLGCSHLVEALYLVCGERGFFYTPKTRGIVEQCCTICSILYOLENYCN 107  
 |||||||  
 DB 1 FVNQHLGCSHLVEALYLVCGERGFFYTPKT-GIVEQCCTGVCSLYOLENYCN 51

## RESULT 3

INS\_CERAE STANDARD; PRT: 110 AA.  
 AC P30407; P01309;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE Insulin precursor.  
 GN INS.  
 OS Cercopithecus aethiops (Green monkey) (Grivet).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;

OC Cercopithecinae; Cercopithecus.  
 OX NCBI\_TaxID=9534;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92219953; PubMed=1560757;  
 RA Seino S., Bell G.I., Li W.;  
 RT \*Sequences of primate insulin genes support the hypothesis of a  
 RT slower rate of molecular evolution in humans and apes than in  
 RT monkeys.\*;  
 RL J. Biol. Evol. 9:193-203(1992).  
 RN [2]  
 RP SEQUENCE OF 57-87.  
 RX MEDLINE=72258016; PubMed=4626369;  
 RA Peterson J.D., Nehrlich S., Oyer P.E., Steiner D.F.;  
 RT \*Determination of the amino acid sequence of the monkey, sheep, and  
 RT dog proinsulin C-peptides by a semi-micro Edman degradation  
 RT procedure.\*;  
 RL J. Biol. Chem. 247:4866-4871(1972).  
 CC \*FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
 CC \*SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
 CC DISULFIDE BONDS.  
 CC \*SUBCELLULAR LOCATION: Secreted.  
 CC \*SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 CC \*THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: X61092; CAA43405.1;  
 DR PIR: B42179; B42179.  
 DR HSSP: P01308; IAI0.  
 DR InterPro: IPR0004825; Ins/IGF/relax.  
 DR SMART: PF0049; Insulin; 1.  
 DR PROSITE: PS00262; IIGF; 1.  
 KW Insulin family; Hormone; Glucose metabolism; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 54 INSULIN B CHAIN.  
 FT PROPEP 57 87 C-PEPTIDE.  
 FT CHAIN 90 110 INSULIN A CHAIN.  
 FT DISULFID 31 96 INTERCHAIN.  
 FT DISULFID 43 104 INTERCHAIN.  
 FT DISULFID 95 100  
 SQ SEQUENCE 110 AA: 12019 MW: 9541548E7H247F4 CR0E4;

Query Match 46.5%; Score 273; DB 1; Length 110;  
 Best Local Similarity 60.2%; Pred. No. 4.1e-22;  
 Matches 53; Conservative 0; Mismatches 1; Indels 34; Gaps 1;

QY 54 PFVNQHLGCSHLVEALYLVCGERGFFYTPKT-----85  
 |||||||  
 DB 23 PAFVNQHLGCSHLVEALYLVCGERGFFYTPKTREAEADPQVGQVGLGGAGSLQPLAL 82

QY 86 -----RGIVEQCCTICSILYOLENYCN 107  
 |||||||

DB 83 EGSIQKRGIVEQCCTICSILYOLENYCN 110

## RESULT 4

INS\_MACFA STANDARD; PRT: 110 AA.  
 ID INS\_MACFA  
 AC P30406; P01309;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE Insulin precursor.

GN INS.  
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Cercopithecoidea;  
OC Cercopithecoidea; Macaca.  
OX NCBI\_TaxID=9541;  
RN [1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=63080474; PubMed=6184262;  
RA Wetekam W., Groneberg J., Leineweber M., Wengenmayer F.,  
RA Winnacker E.-L.;  
RT "The nucleotide sequence of cDNA coding for preproinsulin from the  
RT primate Macaca fascicularis.";  
RL Gene 19:179-183(1982).  
CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
CC DISULFIDE BONDS.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: J00336; AAA36849.1;  
CC PIR: JQ0178; JQ0178.  
DR HSSP: P01308; LA10.  
DR InterPro: IPR004825; Ins/IGF/relax.  
DR Pfam: PF000049; Insulin: 1.  
DR PRINTS: PR00277; INSULINB.  
DR SMART: SM00078; IIGF: 1.  
DR PROSITE: PS00262; INSULIN: 1.  
KW Insulin family; Hormone; Glucose metabolism; Signal.  
FT SIGNAL 1 24  
FT CHAIN 25 54 INSULIN B CHAIN.  
FT PROPEP 57 87 C PEPTIDE.  
FT CHAIN 90 110 INSULIN A CHAIN.  
FT DISULFID 31 96 INSULIN A CHAIN.  
FT DISULFID 43 109 INTERCHAIN.  
FT DISULFID 95 100 INTERCHAIN.  
SQ SEQUENCE 110 AA; 11991 MW; 83C6E33A80A420F9 CRC64;  
Query Match 46.5%; Score 273; DB 1; Length 110;  
Best Local Similarity 60.2%; Prod. No. 4,1e-22;  
Matches 53; Conservative 0; Mismatches 1; Indels 34; Gaps 1;  
Qy 54 PFVNQHLGSHLVEALYLVCGERGFFYTPKT----- 85  
Db 23 PAFVNQHLGSHLVEALYLVCGERGFFYTPKTREAAEDPQGVGLGGPGAGSLQPLAL 82  
Qy 86 -----RGIVEQCTCSLSLYQLENYCN 107  
Db 83 EGSLOKRGIVEQCTCSLSLYQLENYCN 110  
RESULT 5  
INS\_ACOA  
ID INS\_ACOA STANDARD; PRT; 51 AA.  
AC P01324;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Insulin.  
GN INS.  
OS Acomys cahirinus (Egyptian spiny mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Rodentia; Muridae; Murinae; Acomys.  
OX NCBI\_TaxID=10066;  
RN [1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=72189454; PubMed=5028210;  
RA Buentzli H.F., Humbel R.F.;  
RT "Isolation and partial structural analysis of insulin from mouse (Mus  
RT musculus) and spiny mouse (Acomys cahirinus)."  
RL Hoppe-Sevler's 2. Physiol. Chem. 353:444-450(1972).  
CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
CC DISULFIDE BONDS.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
CC PIR: A01591; INMSSP.  
DR HSSP: P01308; ILYM.  
DR InterPro: IPR004825; Ins/IGF/relax.  
DR SMART: SM00078; IIGF: 1.  
DR PROSITE: PS00262; INSULIN: 1.  
KW Insulin family; Hormone; Glucose metabolism.  
FT CHAIN 1 30 INSULIN B CHAIN.  
FT NON-CONS 30 31  
FT CHAIN 31 51 INSULIN A CHAIN.  
FT DISULFID 7 37 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 19 50 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 36 41 BY SIMILARITY.  
SQ SEQUENCE 51 AA; 5768 MW; 92BD88629047D3D CRC64;  
Query Match 45.7%; Score 268.5; DB 1; Length 51;  
Best Local Similarity 92.3%; Prod. No. 5,5e-22;  
Matches 48; Conservative 3; Mismatches 0; Indels 1; Gaps 1;  
Qy 56 PFVNQHLGSHLVEALYLVCGERGFFYTPKT----- 107  
Db 1 FVBQHLGSHLVEALYLVCGERGFFYTPKS-GIVDQCTCSLSLYQLENYCN 51  
RESULT 6  
INS\_HUMAN  
ID INS\_HUMAN STANDARD; PRT; 110 AA.  
AC P01308;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Insulin precursor.  
GN INS.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominoidea.  
OX NCBI\_TaxID=9606;  
RN [1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=80120725; PubMed=6243748;  
RA Bell G.I., Pictet R.L., Rutter W.J., Cordell B., Tischer E.,  
RA Goodman H.M.;  
RT "Sequence of the human insulin gene.";  
RL Nature 284:26-32(1980).  
RN [2]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=80236313; PubMed=6248962;  
RA Ulrich A., Dull T.J., Gray A., Brosius J., Sures I.;  
RT "Genetic variation in the human insulin gene.";  
RL Science 209:612-615(1980).  
RN [3]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=80054779; PubMed=503234;  
RA Bell G.I., Swain W.F., Pictet R.L., Cordell B., Goodman H.M.,  
RA Rutter W.J.;  
RT "Nucleotide sequence of a cDNA clone encoding human preproinsulin.";  
RL Nature 282:525-527(1979).

- RP SEQUENCE FROM N.A.  
RX MEDLINE-80147417; PubMed-6927840;  
RA Sures I., Goeddel D.V., Gray A., Ullrich A.;  
RT "Nucleotide sequence of human preproinsulin complementary DNA.";  
RL Science 208:57-59(1980).  
RN [5]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-93364428; PubMed-8358440;  
RA Lucassen A.M., Bell J.I., Julier C., Lathrop M.;  
RT "Susceptibility to insulin dependent diabetes mellitus maps to a 4.1  
RT kb segment of DNA spanning the insulin gene and associated VNTR.";  
RL Nat. Genet. 4:305-310(1993).  
RN [6]  
RP SEQUENCE FROM N.A.  
RC TISSUE-Pancreas;  
RX MEDLINE-22388257; PubMed-12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Narusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny K.C., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [7]  
RP SEQUENCE OF 1-59 FROM N.A.  
RC TISSUE-Blood;  
RX Pajardy I.I., Weill J.J., Stuckens C.C., Danze P.M.P.;  
RA "Description of a novel RFLP diallelic polymorphism (-127 BsgI C/G)  
RT within the 5' region of insulin gene.";  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
RN [8]  
RP SEQUENCE OF 75-54 AND 90-110.  
RA Nicol D.S.H.W., Smith L.F.;  
RT "Amino-acid sequence of human insulin.";  
RL Nature 187:483-485(1960).  
RN [9]  
RP SEQUENCE OF 57-87.  
RX MEDLINE-71116410; PubMed-5101771;  
RA Over P.E., Cho S., Peterson J.D., Steiner D.F.;  
RT "Studies on human proinsulin. Isolation and amino acid sequence of  
RT the human pancreatic C-peptide.";  
RL J. Biol. Chem. 246:1375-1386(1971).  
RN [10]  
RP SEQUENCE OF 57-87.  
RX MEDLINE-71257722; PubMed-5560404;  
RA Ko A., Smyth D.G., Markussen J., Sundby F.;  
RT "The amino acid sequence of the C-peptide of human proinsulin.";  
RL Eur. J. Biochem. 20:190-199(1971).  
RN [11]  
RP SYNTHESIS.  
RX MEDLINE-75077277; PubMed-4443293;  
RA Steiber P., Kamber B., Hartmann A., Joehl A., Riniker B., Rittel W.;  
RT "Total synthesis of human insulin under directed formation of the  
RT disulfide bonds.";  
RL Helv. Chim. Acta 57:2617-2621(1974).  
RN [12]  
RP SYNTHESIS OF 57-87.  
RX MEDLINE-75040007; PubMed-4803504;  
RA Naithani V.K.;  
RT "Studies on polypeptides, IV. The synthesis of C-peptide of human  
RT proinsulin.";  
RL Hoppe-Seyler's Z. Physiol. Chem. 354:659-672(1973).  
RN [13]  
RP SYNTHESIS OF 65-69 AND 70-73.  
RX MEDLINE-73161263; PubMed-4698555;  
RA Geiger R., Volk A.;  
RT "Synthesis of peptides with the properties of human proinsulin C  
RT peptides (hC peptide). 3. Synthesis of the sequences 14-17 and 9-13  
RL of human proinsulin C peptides.";  
RN Chem. Ber. 106:199-205(1973).  
RP SYNTHESIS OF 84-87.  
RX MEDLINE-73161261; PubMed-4698553;  
RA Geiger R., Jaeger G., Keonig W., Treuth G.;  
RT "Synthesis of peptides with the properties of human proinsulin C  
RT peptides (hC peptide). I. Scheme for the synthesis and preparation of  
RL the sequence 28-31 of human proinsulin C peptide.";  
RN Chem. Ber. 106:188-192(1973).  
RP VARIANT LOS ANGELES SER-48.  
RX MEDLINE-84016053; PubMed-6312455;  
RA Haneda M., Chan S.J., Kwok S.C.M., Rubenstein A.H., Steiner D.F.;  
RT "Studies on mutant human insulin genes: identification and sequence  
RT analysis of a gene encoding [Ser824]insulin.";  
RL Proc. Natl. Acad. Sci. U.S.A. 80:6366-6370(1983).  
RN [16]  
RP VARIANTS LOS ANGELES SER-48 AND CHICAGO LEU-49.  
RX MEDLINE-84170233; PubMed-6424111;  
RA Shoelson S., Fickova M., Haneda M., Nahum A., Musso G., Kaiser E.T.,  
RA Rubenstein A.H., Tager H.;  
RT "Identification of a mutant human insulin predicted to contain a  
RT serine-for-phenylalanine substitution.";  
RL Proc. Natl. Acad. Sci. U.S.A. 80:7390-7394(1983).  
RN [17]  
RP VARIANT PROVIDENCE ASP-34.  
RX MEDLINE-87175640; PubMed-3477744;  
RA Chan S.J., Seino S., Gruppas P.A., Schwartz R., Steiner D.F.;  
RT "A mutation in the B chain coding region is associated with impaired  
RT proinsulin conversion in a family with hyperproinsulinemia.";  
RL Proc. Natl. Acad. Sci. U.S.A. 4:2194-2197(1987).  
RN [18]  
RP VARIANT WAKAYAMA LEU-92.  
RX MEDLINE-87058122; PubMed-3511;  
RA Sakura H., Iwamoto Y., Sakuma Y., Kuzuya T., Hirata H.;  
RT "Structurally abnormal insulin in a diabetic patient. Characterization  
RT of the mutant insulin A3 (Val-->Leu) isolated from the pancreas.";  
RL J. Clin. Invest. 78:1666-1672(1986).  
RN [19]  
RP VARIANT HIS-89.  
RX MEDLINE-90317021; PubMed-2796279;  
RA Barbetti F., Raben N., Katakaki T., Cama A., Accilli D., Gabbay K.H.,  
RA Merenich J.A., Taylor S.I., Roth J.;  
RT "Two unrelated patients with familial hyperproinsulinemia due to a  
RT mutation substituting histidine for arginine at position 65 in the  
RT proinsulin molecule: identification of the mutation by direct  
RT sequencing of genomic deoxyribonucleic acid amplified by polymerase  
RT chain reaction.";  
RL J. Clin. Endocrinol. Metab. 71:164-169(1990).  
RN [20]  
RP VARIANT HIS-89.  
RX MEDLINE-85261996; PubMed-4019786;  
RA Shibasaki Y., Kawakami T., Kanazawa Y., Akanuma Y., Takaku F.;  
RT "Posttranslational cleavage of proinsulin is blocked by a point  
RT mutation in familial hyperproinsulinemia.";  
RL J. Clin. Invest. 76:378-387(1985).  
RN [21]  
RP VARIANT KYOTO LEU-89.  
RX MEDLINE-92291307; PubMed-1601997;  
RA Yano H., Kitano N., Morimoto M., Polonsky K.S., Imura H., Seino Y.;  
RT "A novel point mutation in the human insulin gene giving rise to  
RT hyperproinsulinemia (proinsulin Kyoto).";  
RL J. Clin. Invest. 89:1902-1907(1992).

RN [22]  
 RP STRUCTURE BY NMR.  
 RX MEDLINE-91104966; PubMed-2271664;  
 RA Hua Q.-X., Weiss M.A.;  
 RT "Toward the solution structure of human insulin: sequential 2D 1H NMR  
 assignment of a des-pentapeptide analogue and comparison with crystal  
 structure.";  
 RL Biochemistry 29:10545-10555(1990).  
 RN [23]  
 RP STRUCTURE BY NMR.  
 RX MEDLINE-91242467; PubMed-2036420;  
 RA Hua Q.-X., Weiss M.A.;  
 RT "Comparative 2D NMR studies of human insulin and des-pentapeptide  
 insulin: sequential resonance assignment and implications for protein  
 dynamics and receptor recognition.";  
 RL Biochemistry 30:5505-5515(1991).  
 RN [24]  
 RP STRUCTURE BY NMR.  
 RX MEDLINE-91265527; PubMed-1646635;  
 RA Hua Q.-X., Weiss M.A.;  
 RT "Two-dimensional NMR studies of Des-(B26-B30)-insulin: sequence-  
 specific resonance assignments and effects of solvent composition.";  
 RL Biochim. Biophys. Acta 1078:101-110(1991).  
 Query Match 45.5%; Score 267; DB 1; Length 110;  
 Best Local Similarity 60.5%; Pred. No. 1.8e-21;  
 Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;  
 QY 56 FVNOHLCGSHLVEALYLVCGERGFFYTPKT----- 85  
 DB 25 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVQLVGQGLGGPGASGLPLALEG 84  
 QY 86 ----RGIVEQCCTSLCYOLENYCN 107  
 DB 85 SLOKRGIVEQCCTSLCYOLENYCN 110  
 RESULT 7  
 INS\_PANTR STANDARD; PRT: 110 AA.  
 ID INS\_PANTR  
 AC P30410;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Insulin precursor.  
 GN INS.  
 OS Pan troglodytes (Chimpanzee).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
 OX NCBI\_TaxID=9598;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-92219553; PubMed=1560757;  
 RA Seleno S., Bell G.I., Li W.;  
 RT "Sequences of primate insulin genes support the hypothesis of a  
 slower rate of molecular evolution in humans and apes than in  
 monkeys.";  
 RL Mol. Biol. Evol. 9:193-203(1992).  
 CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
 INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
 FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
 CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
 CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
 DISULFIDE BONDS.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC EMBL; X61089; CAA43403.1; -;  
 DR PIR; A42179; A42179.  
 DR PDB; LEFE; 29-MAR-00.  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1.  
 DR PRINTS: PR00277; INSULINB.  
 DR SMART: SM00078; IIGF; 1.  
 DR PROSITE: PS00262; INSULIN;  
 KW Insulin family; Hormone; Glucose metabolism; Signal; 3D-structure.  
 FT SIGNAL 1 24  
 FT CHAIN 25 54 INSULIN B CHAIN.  
 FT PROPEP 57 87 C PEPTIDE.  
 FT CHAIN 90 110 INSULIN A CHAIN.  
 FT DISULFID 31 96 INTERCHAIN.  
 FT DISULFID 43 109 INTERCHAIN.  
 FT DISULFID 95 100  
 SQ SEQUENCE 110 AA; 12025 MW; 41EB8DF79837CEFS CRC64;  
 Query Match 45.5%; Score 267; DB 1; Length 110;  
 Best Local Similarity 60.5%; Pred. No. 1.8e-21;  
 Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;  
 QY 56 FVNOHLCGSHLVEALYLVCGERGFFYTPKT----- 85  
 DB 25 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVQLVGQGLGGPGASGLPLALEG 84  
 QY 86 ----RGIVEQCCTSLCYOLENYCN 107  
 DB 85 SLOKRGIVEQCCTSLCYOLENYCN 110  
 RESULT 8  
 INS\_SPETR STANDARD; PRT: 110 AA.  
 ID INS\_SPETR  
 AC Q91X13.  
 DT 15-SEP-2003 (Rel. 42, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Insulin precursor.  
 GN INS.  
 OS Spermophilus tridecemlineatus (Thirteen-lined ground squirrel).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;  
 OC Spermophilus.  
 OX NCBI\_TaxID=43179;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreas;  
 RA Tredrea M.M., Buck M.J., Gubaniyogi J., Squire T.L., Andrews M.T.;  
 RT "Regulation of PDK4 expression in a hibernating mammal";  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
 INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
 FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
 CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
 CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
 DISULFIDE BONDS.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC EMBL; AV038604; AAK72558.1; -;  
 DR HSSP; P01308; ILNP.  
 DR InterPro: IPR004825; Ins/IGF/relax.

```
DR pfam: PF00049; Insulin; 1.  
DR PRINTS: PR00277; INSULIN.  
DR SMART: SM00078; IIGF; 1.  
DR PROSITE: PS00262; INSULIN; 1.  
KW Insulin family; Hormone; Glucose metabolism; Signal.  
FT SIGNAL 1 24 BY SIMILARITY.  
FT CHAIN 25 54 INSULIN B CHAIN.  
FT PROPEP 57 87 C PEPTIDE.  
FT CHAIN 90 110 INSULIN A CHAIN.  
FT DISULFID 31 96 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 43 109 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 95 100 BY SIMILARITY.  
SQ SEQUENCE 110 AA; 12004 MW; 4511768D6122BEE5 CRC64;  
  
Query Match 45.3%; Score 266; DB 1; Length 110;  
Best Local Similarity 57.4%; Pred. No. 2.2e-21;  
Matches 54; Conservative 1; Mismatches 3; Indels 36; Gaps 2;  
  
QY 50 LGTGP--RFVNHLCGSHLVYALVLCGERGFYTPKTRGIVEOCCTSGSLYLENYCN 107  
DB 17 LGPDPAQAFVNHLCGSHLVYALVLCGERGFYTPKSRREVVEEQGGQVFLGGPGAGL 76  
QY 85 -----RGIVEOCCTSGSLYLENYCN 107  
DB 77 PQLPALENALQKRGIVEOCCTSGSLYLENYCN 110  
  
RESULT 9  
INS_BALBO STANDARD; PRT; 51 AA.  
AC P01314;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Insulin.  
GN INS.  
OS Balaenoptera borealis (Sei whale).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;  
OC Balaenopteridae; Balaenoptera.  
OX NCBI_TaxID=9768;  
RN [1]  
RP SEQUENCE.  
RA Ishihara Y., Saito T., Ito Y., Fujino M.;  
RT "Structure of sperm- and sei-whale insulins and their breakdown by  
RL Nature 181:1468-1469(1958).  
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
CC DISULFIDE BONDS.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
DR PIR: A01582; INWHIS.  
DR HSSP: P01317; IAPH.  
DR SMART: PR004825; Ins/IGF/relax.  
DR PROSITE: PS00078; IIGF; 1.  
KW Insulin family; Hormone; Glucose metabolism.  
FT CHAIN 1 30 INSULIN B CHAIN.  
FT NON_CONS 30 31  
FT CHAIN 31 51 INSULIN A CHAIN.  
FT DISULFID 7 37 INTERCHAIN.  
FT DISULFID 19 50  
FT DISULFID 36 41  
SQ SEQUENCE 51 AA; 5723 MW; 9007B50E400A7DDD CRC64;  
  
Query Match 44.9%; Score 263.5; DB 1; Length 51;  
Best Local Similarity 92.3%; Pred. No. 1.9e-21;  
Matches 48; Conservative 0; Mismatches 3; Indels 1; Gaps 1;  
  
pfam: PF00049; Insulin; 1.  
PRINTS: PR00277; INSULIN.  
SMART: SM00078; IIGF; 1.  
PROSITE: PS00262; INSULIN; 1.  
Insulin family; Hormone; Glucose metabolism; Signal.  
SIGNAL 1 24 BY SIMILARITY.  
CHAIN 25 54 INSULIN B CHAIN.  
PROPEP 57 87 C PEPTIDE.  
CHAIN 90 110 INSULIN A CHAIN.  
DISULFID 31 96 INTERCHAIN (BY SIMILARITY).  
DISULFID 43 109 INTERCHAIN (BY SIMILARITY).  
DISULFID 95 100 BY SIMILARITY.  
SEQUENCE 110 AA; 12004 MW; 4511768D6122BEE5 CRC64;  
  
Query Match 45.3%; Score 266; DB 1; Length 110;  
Best Local Similarity 57.4%; Pred. No. 2.2e-21;  
Matches 54; Conservative 1; Mismatches 3; Indels 36; Gaps 2;  
  
QY 50 LGTGP--RFVNHLCGSHLVYALVLCGERGFYTPKTRGIVEOCCTSGSLYLENYCN 107  
DB 17 LGPDPAQAFVNHLCGSHLVYALVLCGERGFYTPKSRREVVEEQGGQVFLGGPGAGL 76  
QY 85 -----RGIVEOCCTSGSLYLENYCN 107  
DB 77 PQLPALENALQKRGIVEOCCTSGSLYLENYCN 110  
  
RESULT 10  
INS_CAMDR STANDARD; PRT; 51 AA.  
AC P01320;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 01-FEB-1996 (Rel. 33, Last annotation update)  
DE Insulin.  
GN INS.  
OS Camelus dromedarius (Dromedary) (Arabian camel).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.  
OX NCBI_TaxID=9838;  
RN [1]  
RP SEQUENCE.  
RA Danho W.O.;  
RT "The isolation and characterization of insulin of camel (Camelus  
RT dromedarius)".  
PL J. Fac. Med. Baghdad 14:16-28(1972).  
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
CC DISULFIDE BONDS.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
DR PIR: A92782; INCA.  
DR HSSP: P01317; 2INS.  
DR SMART: SM00078; IIGF; 1.  
DR PROSITE: PS00262; INSULIN; 1.  
KW Insulin family; Hormone; Glucose metabolism.  
FT CHAIN 1 30 INSULIN B CHAIN.  
FT NON_CONS 30 31  
FT CHAIN 31 51 INSULIN A CHAIN.  
FT DISULFID 7 37 INTERCHAIN.  
FT DISULFID 19 50  
FT DISULFID 36 41  
SQ SEQUENCE 51 AA; 5693 MW; 901E88BA085A70DD CRC64;  
  
Query Match 44.9%; Score 263.5; DB 1; Length 51;  
Best Local Similarity 90.4%; Pred. No. 1.9e-21;  
Matches 47; Conservative 1; Mismatches 3; Indels 1; Gaps 1;  
  
QY 56 FVNHLCGSHLVYALVLCGERGFYTPKTRGIVEOCCTSGSLYLENYCN 107  
DB 1 FVNHLCGSHLVYALVLCGERGFYTPKA-GIVEOCCASVCSLYOLENYCN 51  
  
RESULT 11  
INS_CAPHI STANDARD; PRT; 51 AA.  
AC P01319;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 01-FEB-1996 (Rel. 33, Last annotation update)  
DE Insulin.  
GN INS.  
OS Capra hircus (Goat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Caprinae; Capra.  
OX NCBI_TaxID=9925;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=66160119; PubMed=5949593;  
RA Smith L.F.;
```

```

RT *Species variation in the amino acid sequence of insulin.*
RL Am. J. Med. 40:662-666(1966).
CC -I- INCREASES INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. II
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -I- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR PIR: A01586; INGT.
DR HSSP: P01317; IAPH.
DR InterPro: IPR004825; Ins/IGF/relax.
DR SMART: SM00078; ILGF: 1.
DR PROSITE: PS00262; INSULIN: 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON-CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5692 MW; 9007B50CDB4E7DDD CRC64;

Query Match 44.98; Score 263.5; DB 1; Length 51;
Best Local Similarity 90.48; Pred. No. 1.9e-21;
Matches 47; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 56 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTISCSLYOLENYCN 107
DDB 1 FVNHLCGSHLVEALYLVCGERGFFYTPKA-GIVEQCAGVCSLYOLENYCN 5;

RESULT 12
INS_PIG
ID INS_PIG STANDARD; PRT; 108 AA.
AC P01315; O9TSJ5;
DI 21-JUL-1986 (Rel. 01, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Insulin precursor.
GN INS.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Han X.G., Tsch B.E.;
RT *Complete porcine preproinsulin cDNA sequence.*;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Large white;
RX MEDLINE=22135958; PubMed=12140686;
RA Amarger V., Nguyen M., Laere A.S., Braunschweig M., Nezer C.,
RA Georges M., Andersson L.;
RT *Comparative sequence analysis of the INS-IGF2-H19 gene cluster in
RT pigs.*;
RL Mamm. Genome 13:388-398(2002).
RN [3]
RP SEQUENCE OF 25-108.
RX MEDLINE=68286485; PubMed=5657063;
RA Chance R.E., Ellis R.M., Bromer W.W.;
RT *Porcine proinsulin: Characterization and amino acid sequence.*;
RL Science 161:165-167(1968).
RN [4]
RP REVISION TO 59.
RA Chance R.E.;
RL Submitted (JUL-1970) to the PIR data bank.
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RA Blundell T.L., Dodson G.G., Hodgkin D., Mercola D.;

```

```

RT *Insulin. The structure in the crystal and its reflection in
RT chemistry and biology.*;
RL Adv. Protein Chem. 26:279-402(1972).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
RA Isaacs N.W., Aqarwal R.C.;
RT *Experience with fast Fourier least squares in the refinement of the
RT crystal structure of rhombohedral 2-zinc insulin at 1.5-A
RT resolution.*;
RL Acta Crystallogr. A 34:782-791(1978).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
RX MEDLINE=89099418; PubMed=2905485;
RA Baker E.N., Blundell T.L., Cutfield J.F., Cutfield S.M., Dodson E.J.,
RA Dodson G.G., Crowfoot Hodgkin D.M., Hubbard R.E., Isaacs N.W.,
RA Reynolds C.D., Sakabe K., Sakabe N., Vijayan N.M.;
RT *The structure of 2Zn pig insulin crystals at 1.5-A resolution.*;
RL Philos. Trans. R. Soc. Lond. B, Biol. Sci. 319:369-456(1988).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=92126280; PubMed=1772633;
RA Balschmidt P., Hansen F.B., Dodson E., Dodson G., Korber F.;
RT *Structure of porcine insulin co-crystallized with clupeine Z.*;
RL Acta Crystallogr. B 47:975-986(1991).
RN [9]
RP X-RAY CRYSTALLOGRAPHY.
RX MEDLINE=91222450; PubMed=2025410;
RA Badger J., Harris M.R., Reynolds G.D., Evans A.C., Dodson E.J.,
RA Dodson G.G., North A.C.T.;
RT *Structure of the pig insulin dimer in the cubic crystal.*;
RL Acta Crystallogr. B 47:127-136(1991).
RN [10]
RP X-RAY CRYSTALLOGRAPHY (1.65 ANGSTROMS).
RA Diau J.-S., Wan Z.-L., Chang W.-R., Liang D.-C.;
RT *Structure of monomeric porcine DesB1-B2 desptenapeptide (B26-B40)
RT insulin at 1.65-A resolution.*;
RL Acta Crystallogr. D 53:507-512(1997).
CC -I- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -I- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC -I- DATABASE: NAME-Protein Spotlight;
CC NOTE=Issue 9 of April 2001;
CC WWW="http://www.expasy.org/spotlight/articles/sptlit009.html".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF064555; AAC77920.1; ALT_INIT.
CC EMBL: AY044828; AAL69550.1;
CC PDB: 3INS; 09-JAN-89.
CC PDB: 4INS; 31-JUL-94.
CC PDB: 6INS; 31-JAN-94.
CC PDB: 7INS; 31-JAN-94.
CC PDB: 9INS; 15-OCT-91.
CC PDB: 11ZA; 15-OCT-91.
CC PDB: 11ZB; 15-OCT-91.
CC PDB: 12CI; 29-JAN-96.
CC PDB: 1MPJ; 29-JAN-96.
CC PDB: 3MTH; 29-JAN-96.
CC PDB: 1DEI; 16-JUN-97.
CC PDB: 1SD8; 01-APR-98.
CC PDB: 1WAV; 28-FEB-97.
CC PDB: 1ZEI; 16-FEB-99.

```

DR PDB; 1ZNI; 28-JAN-98.  
 DR PDB; 1ZNI; 28-JAN-98.  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1.  
 DR SMART: SM00078; IIGF; 1.  
 DR PROSITE: PS00262; INSULIN; 1.  
 KW Insulin family; Hormone; Glucose metabolism; Signal; 3D-structure.  
 FT SIGNAL 1 24  
 FT CHAIN 25 54 INSULIN B CHAIN.  
 FT PROPEP 57 85 C PEPTIDE.  
 FT CHAIN 88 108 INSULIN A CHAIN.  
 FT DISULFID 31 94 INTERCHAIN.  
 FT DISULFID 43 107 INTERCHAIN.  
 FT DISULFID 93 98  
 FT HELIX 26 46  
 FT STRAND 48 48  
 FT HELIX 89 94  
 FT HELIX 100 106  
 FT STRAND 107 107  
 SQ SEQUENCE 108 AA; 11671 MW; CB4491B429858EBC CRC64;  
 Query Match 44.8%; Score 263; DB 1; Length 108;  
 Best Local Similarity 60.7%; Pred. No. 4.6e-21;  
 Matches 51; Conservative 0; Mismatches 1; Indels 32; Gaps 1;  
 QY 56 FVNHLCGSHLVEALYLVCGERGFFYTPKT----- 85  
 DB 25 FVNHLCGSHLVEALYLVCGERGFFYTPKRRARENPAQAVELGGGLGQLALAEQGP 84  
 QY 86 --RGIVEQCCTSCSYOLENYCN 107  
 DB 85 QKRGIVEQCCTSCSYOLENYCN 108  
 RESULT 13  
 ID INS\_RAB1 STANDARD; PRT; 110 AA.  
 AC P01311;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Insulin precursor.  
 GN INS.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-New Zealand white; IISU-E-Pancreas;  
 RX MEDLINE=94179230; PubMed=8132571;  
 RA Devaskar S.U.; Giddings S.J.; Rajakumar P.A.; Carnaghi L.R.;  
 RA Menon R.K.; Zahm D.S.;  
 RT "Insulin gene expression and insulin synthesis in mammalian neuronal  
 cells.";  
 RN J. Biol. Chem. 269:8445-8454(1994).  
 RL [2]  
 RN SEQUENCE OF 25-54 AND 90-110.  
 RP MEDLINE=66160119; PubMed=594593;  
 RA Smith L.F.;  
 RT "Species variation in the amino acid sequence of insulin.";  
 RL Am. J. Med. 40:662-666(1966).  
 RN [3]  
 RP SEQUENCE OF 56-110 FROM N.A.  
 RA Giddings S.J.; Carnaghi L.R.; Devaskar S.U.;  
 RA Submitted (APR-1991) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
 CC -1- SUBUNIT: HETEROODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
 CC DISULFIDE BONDS.  
 CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; U03610; AAA19033.1;  
 CC EMBL; M61153; AAA17540.1;  
 CC PIR: A53438; INRB.  
 CC HSSP: P0108; ILYM.  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1.  
 DR SMART: SM00078; IIGF; 1.  
 DR PROSITE: PS00262; INSULIN; 1.  
 KW Insulin family; Hormone; Glucose metabolism; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 54 INSULIN B CHAIN.  
 FT PROPEP 57 87 C PEPTIDE.  
 FT CHAIN 90 110 INSULIN A CHAIN.  
 FT DISULFID 31 96 INTERCHAIN.  
 FT DISULFID 43 109 INTERCHAIN.  
 FT DISULFID 95 100  
 FT CONFLICT 83 83 E -> Y (IN REF. 3).  
 SQ SEQUENCE 110 AA; 11638 MW; 82D2975885D7FA8 CRC64;  
 Query Match 44.8%; Score 263; DB 1; Length 110;  
 Best Local Similarity 59.3%; Pred. No. 4.6e-21;  
 Matches 51; Conservative 1; Mismatches 0; Indels 34; Gaps 1;  
 QY 56 FVNHLCGSHLVEALYLVCGERGFFYTPKT----- 85  
 DB 25 FVNHLCGSHLVEALYLVCGERGFFYTPKRRVEELVQAEIIGCGKAGLQPSALEL 84  
 QY 86 ----RGIVEQCCTSCSYOLENYCN 107  
 DB 85 ALQKRGIVEQCCTSCSYOLENYCN 110  
 RESULT 14  
 ID INS\_FELCA STANDARD; PRT; 51 AA.  
 AC P03106;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Insulin.  
 GN INS.  
 OS Felis silvestris catus (Cat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.  
 OX NCBI\_TaxID=9685;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=86214076; PubMed=3518635;  
 RA Halliden G.; Gavellin G.; Mutt V.; Joernvall H.;  
 RT "Characterization of cat insulin."  
 RL Arch. Biochem. Biophys. 247:20-27(1986).  
 CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
 CC -1- SUBUNIT: HETEROODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
 CC DISULFIDE BONDS.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 CC PIR: A01588; INCT.  
 DR HSSP: P01317; IAPH.  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR PRINTS: PR00277; INSULINB.

[illegible]

```

RESULT 15
INS_CANFA          STANDARD:          PRT: 110 AA.
ID      INS_CANFA
AC      P01321;
DT      21-JUL-1986 (Rel. 01, Created)
DT      21-JUL-1986 (Rel. 01, Last sequence update)
DT      01-FEB-1996 (Rel. 33, Last annotation update)
DE      Insulin precursor.
DE      INS.
GN      Canis familiaris (Dog).
OS      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX      NCBI_TaxID=9615;
[1]
SEQUENCE FROM N.A.
RP      MEDLINE=83109071; PubMed=6296142;
RP      Kwock S.C.M., Chan S.J., Steiner D.F.;
RA      *Cloning and nucleotide sequence analysis of the dog insulin gene.
RT      Coded amino acid sequence of canine preproinsulin predicts an
RT      additional C-peptide fragment.;
RL      J. Biol. Chem. 258:2357-2363(1983).
[2]
SEQUENCE OF 25-54 AND 90-110.
RP      MEDLINE=66160119; PubMed=5949593;
RP      Smith L.F.;
RA      *Species variation in the amino acid sequence of insulin.;
RL      Am. J. Med. 40:662-666(1966).
CC      -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC      INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC      FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC      CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC      -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC      DISULFIDE BONDS.
CC      -1- SUBCELLULAR LOCATION: Secreted.
CC      -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
CC      EMBL: V00179; CAA23475.1;
CC      DR      PIR: A92413; IPDG.
CC      DR      HSSP: P01317; IAPH.
CC      DR      InterPro: IPR004825; Ins/IGF/relax.
CC      DR      Pfam: PF000049; Insulin; 1.
CC      DR      SMART: SM00078; IIGF; 1.
CC      DR      PROSITE: PS00262; INSULIN; 1.
CC      DR      Insulin family; Hormone; Glucose metabolism; Signal.
CC      FT      SIGNAL 1 24

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: September 16, 2003, 12:34:00 : Search time 25.8132 Seconds  
(without alignments)  
1669.670 Million cell updates/sec

**Title:** us-09-423-100-6

**Perfect score:** 587

**Sequence:** 1 MFPTPLRLFDNAWLAHRR.....IVEQCCTSCSYQLENVCN 107

**Scoring table:** BLOSUM62

Gapop 10.0 , Gapext 0.5

**Searched:** 830525 seqs, 258052604 residues

**Total number of hits satisfying chosen parameters:** 830525

**Minimum DB seq length:** 0

**Maximum DB seq length:** 2000000000

**Post-processing:** Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

**Database :**

- SPTREMBL23:\*
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phase:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*
  - 15: sp\_rvirus:\*
  - 16: sp\_bacteriap:\*
  - 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	267	45.5	110	6 Q8HXV2	Q8hxv2 pongo pygma
2	266	45.3	110	11 Q91XI3	Q91xi3 spermophilu
3	251	42.8	110	6 Q8WNW6	Q8wnw6 felis silve
4	249	42.4	217	6 Q8WNE0	Q8wne0 ateles geof
5	228	38.8	245	4 Q14644	Q14644 homo sapien
6	219.5	37.4	106	13 Q91807	Q918q? rana sapien
7	213	36.3	212	6 Q07368	Q07368 macaca mula
8	213	36.3	217	6 Q07367	Q07367 macaca mula
9	201.5	34.3	110	13 Q98TA8	Q98ta8 pantodon bu
10	201	34.2	217	6 Q07369	Q07369 macaca mula
11	197	33.6	217	4 Q14407	Q14407 homo sapien
12	195.5	33.3	108	13 Q9DD55	Q9dde5 brachydanio
13	195.5	33.3	108	13 Q90ZN4	Q90zn4 catla catla
14	195	33.2	111	13 Q98TB0	Q98tb0 chitala chi
15	195	33.2	217	6 Q8WND9	Q8wnd9 ateles geof
16	193.5	33.0	110	13 Q90ZY1	Q90zy1 hiodon alos

17	191.5	32.6	111	13 Q98TA7	Q98ta7 osteoglossu
18	189.5	32.3	108	13 Q98TB1	Q98tb1 catostomus
19	187.5	31.9	87	13 Q98TA9	Q98ta9 graithonemus
20	185.5	31.6	91	13 Q98TB2	Q98tb2 ambloplites
21	171.5	29.2	217	6 Q8MI74	Q8mi74 callithrix
22	161	27.4	216	11 Q70615	Q70615 spalax leuc
23	159.5	27.2	52	6 Q9TV91	Q9tv91 equus caball
24	159.5	27.2	216	6 QMI73	Q8mi73 delphinus d
25	159.5	27.2	216	6 Q8HYE5	Q8hye5 alluoropoda
26	156	26.6	216	11 Q9R2C3	Q9r2c3 mus musculus
27	154	26.2	216	11 Q9JRM4	Q9jrm4 cavia porce
28	152	25.9	178	6 Q95MJ5	Q95mj5 tarsius ban
29	149.5	25.5	204	6 Q95205	Q95205 ovis aries
30	149	25.4	132	13 Q8AV14	Q8av14 petromyzon
31	148.5	25.3	159	13 Q93607	Q93607 paralichthy
32	148.5	25.3	182	13 Q73720	Q73720 oreochromis
33	148.5	25.3	182	13 Q42889	Q42889 oreochromis
34	148.5	25.3	182	13 P79824	P79824 oreochromis
35	148.5	25.3	185	13 Q57436	Q57436 paralichthy
36	148.5	25.3	186	13 Q93527	Q93527 paralichthy
37	148	25.2	210	13 Q91443	Q91443 squalus aca
38	147.5	25.1	116	13 Q91161	Q91161 oncorhynchus
39	147.5	25.1	117	13 Q91476	Q91476 salmo salar
40	147.5	25.1	145	13 Q91475	Q91475 salmo salar
41	147.5	25.1	149	13 Q91231	Q91231 oncorhynchus
42	147.5	25.1	155	13 Q91162	Q91162 oncorhynchus
43	147.5	25.1	161	13 Q91230	Q91230 oncorhynchus
44	147.5	25.1	188	13 Q81268	Q81268 oncorhynchus
45	147.5	25.1	188	13 Q91965	Q91965 oncorhynchus

#### ALIGNMENTS

RESULT 1

Q8HXV2 ID Q8HXV2 PRELIMINARY; PRT; 110 AA.  
AC Q8HXV2;  
DT 01-MAR-2003 (Tremblrel. 23, Created)  
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Insulin precursor.  
GN INS.  
OS Pongo pygmaeus (orangutan)  
OC Eukaryota; Metazoa; Chordata; Gracilata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catherhini; Eumetardae; Pando.  
OX NCBI\_TaxID=9550;  
RN [1];  
RP SEQUENCE FROM N.A.  
RA Stead J.D.H., Jeffreys A.J.;  
RT "Haplotype diversity at the insulin region."  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AY137503; AAN06937.1;  
SQ SEQUENCE 110 AA; 12038 MW; 22D2B32B94F520F8 CRC64;

Query Match 45.5%; Score 267; DB 6; Length 110;  
Best Loca. Similarity 60.5%; Pred. No. 6.6e-24;  
Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

Qy 56 FVNHLCGSHLVEALYLVCGERGFFYTPKT----- 85  
Db 25 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREARDLQGVGLGCGPGAGSLOPLALEG 84  
Qy 86 ----RGIVEQCCTSCSYQLENVCN 107  
Db 85 SLQKRGIVEQCCTSCSYQLENVCN 110

RESULT 2

Q91XI3 ID Q91XI3 PRELIMINARY; PRT; 110 AA.  
AC Q91XI3;  
DT 01-DEC-2001 (Tremblrel. 19, Created)

DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Insulin.  
 OS *Spermophilus tridecemlineatus* (Thirteen-lined ground squirrel).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;  
 OC *Spermophilus*.  
 OX NCBI\_TaxID=43179;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreas;  
 RT Tredrea M.M., Buck M.J., Guhaniyogi J., Squire T.L., Andrews M.T.;  
 RA "Regulation of PK4 expression in a hibernating mammal.";  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 DR EMBL: AF038604; AAK72558.1;  
 DR HSSP: P01308; ILNP;  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1.  
 DR PRINTS: PR00277; INSULINB.  
 DR SMART: SMO078; IIGF; 1.  
 DR PROSITE: P10262; INSULIN; 1.  
 SQ SEQUENCE 110 AA; 12069 MW; 4511768D66228EE5 CRC64;

Query Match 45.3%; Score 266; DB 11; Length 110;  
 Best Local Similarity 57.4%; Pred. No. 8.7e-24;  
 Matches 54; Conservative 1; Mismatches 3; Indels 36; Gaps 2;

QY 50 LCTGTP-*RFVNHLCGSHLVEALYLVCGERGFFYTPKT*----- 85  
 DB 17 LQPDPAQAFVNHLCGSHLVEALYLVCGERGFFYTPKSRHVEEQGGVGLGGPGAGL 76

QY 86 -----RGIVEQCCTSI~~CS~~LYOLENYCN 107  
 DB 77 PQLALENALQKRGIVEQCCTSI~~CS~~LYOLENYCN 110

RESULT 3  
 Q8WNW6  
 ID Q8WNW6 PRELIMINARY; PRI: 110 AA.  
 AC Q8WNW6;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2003 (TREMBlrel. 20, Last sequence update)  
 DE Preproinsulin.  
 OS *Felis silvestris catus* (Cat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.  
 OX NCBI\_TaxID=9685;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreas;  
 RA Okamoto S., Morimatsu M.;  
 RT "Cat insulin.";  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 DR EMBL: AB043535; BAB84110.1;  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1.  
 DR PRINTS: PR00277; INSULINB.  
 DR SMART: SMO078; IIGF; 1.  
 DR PROSITE: PS00262; INSULIN; 1.  
 SQ SEQUENCE 110 AA; 12069 MW; 95F86E170C7BECA4 CRC64;

Query Match 42.8%; Score 251; DB 6; Length 110;  
 Best Local Similarity 55.8%; Pred. No. 5.3e-22;  
 Matches 48; Conservative 2; Mismatches 22; Indels 34; Gaps 1;

QY 56 FVNHLCGSHLVEALYLVCGERGFFYTPKT----- 85  
 DB 25 FVNHLCGSHLVEALYLVCGERGFFYTPKARPEADLQKDAELGEAPGAGGLQPSALEA 84

QY 86 -----RGIVEQCCTSI~~CS~~LYOLENYCN 107  
 DB 85 PLOKRGIVEQCCTSI~~CS~~LYOLENYCN 110

RESULT 4  
 Q8WNEO  
 ID Q8WNEO PRELIMINARY; PRI: 217 AA.  
 AC Q8WNEO;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Growth hormone.  
 GN GH-N.  
 OS *Ateles geoffroyi* (Black-handed spider monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Ateleinae; Ateles.  
 OX NCBI\_TaxID=9509;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Revol A., Esquivel D., Santiago D., Barrera Saldana H.;  
 RT "Independent duplication of the growth hormone gene in three  
 Anthropoidcan lineages.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF374234; AAL72286.1;  
 DR InterPro: IPR001400; Somatotropin.  
 DR Pfam: PF00103; hormone; 1.  
 DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
 DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
 SQ SEQUENCE 217 AA; 24894 MW; 425829FF41EEAAE6 CRC64;

Query Match 42.4%; Score 249; DB 6; Length 217;  
 Best Local Similarity 97.9%; Pred. No. 2e-21;  
 Matches 47; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 FPTIPLSRFLFDNAKLRAKLHOLAFDTYGFEEAYIPKEQKYSFLONP 49  
 DB 27 FPTIPLSRLLDNLRAKLHOLAFDTYGFEEAYIPKEQKYSFLONP 74

RESULT 5  
 Q14644  
 ID Q14644 PRELIMINARY; PRI: 245 AA.  
 AC Q14644;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Placental growth hormone isoform hGH V4 precursor.  
 GN HGH-V.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Term placenta;  
 RX MEDLINE:98373737; PubMed:9709963;  
 RA Boguszewski C.L., Svensson P.A., Jansson T., Clark R.,  
 RA Carlsson L.M.S., Carlsson B.;  
 RT "Cloning of two novel growth hormone transcripts expressed in human  
 placenta.";  
 RL J. Clin. Endocrinol. Metab. 81:2878-2885(1998).  
 DR EMBL: AF006061; AAB71829.1;  
 DR HSSP: P01241; IA22.  
 DR InterPro: IPR001400; Somatotropin.  
 DR Pfam: PF00103; hormone; 1.  
 DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
 FT SIGNAL  
 SQ SEQUENCE 245 AA; 27101 MW; 14CC7F8CD75D91C8 CRC64;

Query Match 38.8%; Score 228; DB 4; Length 245;

```
Best Local Similarity 91.7%; Pred. No. 7.2e-19;
Matches 44; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 FPTPLSLFONAMLRHRLHQLAFDTYOEEFEAYIPKEQKYSFLQNP 49
   |||||
Db 27 FPTPLSLFONAMLRHRLHQLAFDTYOEEFEAYIPKEQKYSFLQNP 74
   |||||

RESULT 6
QYID Q918Q7 PRELIMINARY: PRT: 106 AA.
AC Q918Q7;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Preproinsulin.
OS Rana pipiens (Northern leopard frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Ranidae; Rana.
OX NCBI_TaxID=8404;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20362507; PubMed=10818274;
RA Irwin D.M., Sivarajah P.;
RT *Proinsulin cDNAs from the leopard frog, Rana pipiens: evolution of
RT proinsulin processing.*;
RL Comp. Biochem. Physiol. 125B:405-410(2000).
CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL: AF227187; AAF87285.1;
DR HSSP: P01315; ISDB.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin.1.
DR PRINTS: PR00277; INSULINB.
DR SMART: SM00078; IIGF.1.
DR PROSITE: PS00262; INSULIN.1.
DR SEQUENCE 106 AA; 12183 MW; 3A870EEC70217F92 CRC64;

Query Match 37.4%; Score 219.5; DB 13; Length 106;
Best Local Similarity 49.4%; Pred. No. 2.8e-18;
Matches 41; Conservative 7; Mismatches 4; Indels 31; Gaps 1;

QY 56 FVNOHLCSHLYEALYLVCGERGFFYTPKTR----- 86
   |||||
Db 24 FQNOYLCGSHLYEALYMGCGRGFFSPRSRDLPOLVNLGQSELDEMVOQSOAFQKR 83
   |||||

QY 87 --GIVEQCCTSCSLYLENYCN 107
   |||||
Db 84 KPG:VEQCCHNTCSLYDLENYCN 106
   |||||

RESULT 7
QYID Q07368 PRELIMINARY: PRT: 212 AA.
AC Q07368;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Somatotropin 2 precursor (Growth hormone 2) (Fragment).
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Placenta;
RX MEDLINE=94008724; PubMed=8404617;
RA Golos T.G., Durning M., Fisher J.M., Fowler P.D.;
RT *Cloning of four growth hormone/chorionic somatomammotropin-related
RT complementary deoxyribonucleic acids differentially expressed during
RT pregnancy in the rhesus monkey placenta.*;
RL Endocrinology 133:1744-1752(1993).
```

```
DR EMBL: L16553; AAA18840.1;
DR HSSP: P01241; IAXI.
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; Hormone.1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
FT NON_TER 1
SQ SEQUENCE 212 AA; 24525 MW; 27BC91106256E6F5 CRC64;

Query Match 36.3%; Score 213; DB 6; Length 212;
Best Local Similarity 78.7%; Pred. No. 3.7e-17;
Matches 37; Conservative 9; Mismatches 1; Indels 0; Gaps 0;

QY 3 PTIPLSLFONAMLRHRLHQLAFDTYOEEFEAYIPKEQKYSFLQNP 49
   |||||
Db 23 PSVPLSLRFLDHAMIQAHRLHQLAFDTYOEEFEAYIPKEKKHSLMENP 69
   |||||

RESULT 8
QYID Q07367 PRELIMINARY: PRT: 217 AA.
AC Q07367;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Chorionic somatomammotropin-1.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Midpregnancy placenta;
RX MEDLINE=94008724; PubMed=8404617;
RA Golos T.G., Durning M., Fisher J.M., Fowler P.D.;
RT *Cloning of four growth hormone/chorionic somatomammotropin-related
RT complementary deoxyribonucleic acids differentially expressed during
RT pregnancy in the rhesus m key placenta.*;
RL Endocrinology 133:1744-1752(1993).
DR EMBL: L16552; AAA18839.1;
DR HSSP: P01241; IAXI.
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; Hormone.1.
DR PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
DR SEQUENCE 217 AA; 24942 MW; FFAA8915131F2BC CRC64;

Query Match 36.3%; Score 213; DB 6; Length 217;
Best Local Similarity 78.7%; Pred. No. 3.8e-17;
Matches 37; Conservative 9; Mismatches 1; Indels 0; Gaps 0;

QY 3 PTIPLSLFONAMLRHRLHQLAFDTYOEEFEAYIPKEQKYSFLQNP 49
   |||||
Db 28 PSVPLSLRFLDHAMIQAHRLHQLAFDTYOEEFEAYIPKEKKHSLMENP 74
   |||||

RESULT 9
QYID Q98TA8 PRELIMINARY: PRT: 110 AA.
AC Q98TA8;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Preproinsulin.
OS Pantodon buchholzi (Butterflyfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;
OC Osteoglossiformes; Pantodontidae; Pantodon.
OX NCBI_TaxID=8276;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21203571; PubMed=11306171;
RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;
```

RT \*Molecular cloning of preproinsulin cDNAs from several  
 RL osteoelossomorphs and a cyprinid.\*;  
 CC Mol. Cell. Endocrinol. 174:51-58(2001).  
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.

DR EMBL: AF199588; AAK28712.1; -;  
 DR HSSP: P01308; 1HIS.  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1.  
 DR SMART: SM00078; IIGF; 1.  
 DR PROSITE: PS00262; INSULIN; 1.  
 SQ SEQUENCE 110 AA; 12324 MW; BDECCD659D872E06 CRC64;

Query Match 34.3%; Score 201.5; DB 13; Length 110;  
 Best Local Similarity 43.5%; Pred. No. 4e-16;  
 Matches 37; Conservative 8; Mismatches 5; Indels 35; Gaps 1;

QY 58 NQHLCGSHLYALVLCGERGFFVTPKT----- 85

DB 26 SOHLCGSHLYDALVMCGEGFFVPTKRDVDPLLGLFSPKSAQENAEDEYPKDQGL 85

QY 86 ---RGIVEQCCTSCSLYQLENYCN 107

DB 86 KVKRGIVEQCCHPCNIFDLQNYCN 110

## RESULT 10

ID Q07369 PRELIMINARY; PRT: 217 AA.

AC Q07369;

DI 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Chorionic somatomotropin-3.

OS Macaca mulatta (Rhesus macaque).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;

OC Cercopithecoidea; Macaca.

OX NCBI\_TaxID=9544;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Midpregnancy placenta;

RX MEDLINE=94008724; PubMed=8404617;

RA Golos T.G., Durning M., Fisher J.M., Fowler P.D.;

RA \*Cloning of four growth hormone/chorionic somatomotropin-related

RT complementary deoxyribonucleic acids differentially expressed during

RT pregnancy in the rhesus monkey placenta.\*;

RL Endocrinology 133:1744-1752(1993).

DR EMBL: L16554; AAA18841.1; -;

DR HSSP: P01241; IAXI.

DR InterPro: IPR001400; Somatotropin.

DR Pfam: PF00103; hormone; 1.

DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.

DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.

SQ SEQUENCE 217 AA; 24874 MW; FIER6AFDBA1B185 CRC64;

Query Match 34.2%; Score 201; DB 6; Length 217;

Best Local Similarity 74.5%; Pred. No. 1e-15;

Matches 35; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

QY 3 TPIPLSLFDNMLRAHLQLAFDTYQEFEEAYIPKEQKYSFLONP 49

DB 28 PSVPLSLRFDNMHAHLQLAFDTYQEFEEAYIPKEKXSLMGNP 74

## RESULT 11

ID Q14407 PRELIMINARY; PRT: 217 AA.

AC Q14407;

DI 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Chorionic somatomotropin CS-2 (Chorionic somatomotropin hormone

DE 2).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89307277; PubMed=2744760;  
 RA Chen E.Y., Liao Y.C., Smith D.H., Barrera-Saldana H.A., Gellinas R.E.,  
 RA Seeburg P.H.;  
 RT \*The human growth hormone locus: nucleotide sequence, biology, and  
 RT evolution.\*;  
 RL Genomics 4:479-497(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91102558; PubMed=1158;  
 RA Vnencak-Jones C.L., Phillips J.A. III;  
 RT \*Hot spots for growth hormone gene deletions in homologous regions  
 RT outside of Alu repeats.\*;  
 RL Science 250:1745-1748(1990).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA Strausberg R.;  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA Strausberg R.;  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL: J01071; AA52553.1; -;  
 DR EMBL: BC022044; AAH22044.1; -;  
 DR EMBL: BC035965; AAH35965.1; -;  
 DR HSSP: P01241; 1A22.  
 DR InterPro: IPR001400; Somatotropin.  
 DR Pfam: PF00103; hormone; 1.  
 DR PROSITE: PR00836; SOMATOTROPIN  
 DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
 DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.

SQ SEQUENCE 217 AA; 24994 MW; 39FAACDDB6B62E951 CRC64;

Query Match 33.6%; Score 197; DB 4; Length 217;

Best Local Similarity 80.0%; Pred. No. 3e-15;

Matches 36; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 4 TPIPLSLFDNMLRAHLQLAFDTYQEFEEAYIPKEQKYSFLON 48

DB 29 TVPLSLRFDNMLRAHLQLAFDTYQEFEEAYIPKEQKYSFLHD 73

## RESULT 12

Q00DE5 PRELIMINARY; PRT: 108 AA.

ID Q00DE5;

DI 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Insulin precursor.

GN INS.

OS Brachydanio rerio (Zebrafish) (Danio rerio).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;

OC Cyprinidae; Danio.

OX NCBI\_TaxID=7955;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99425190; PubMed=10495291;

RA Argenton F., Zecchin E., Bortolussi M.;

RT \*Early appearance of pancreatic hormone-expressing cells in the

RT zebrafish embryo.\*;

RL Mech. Dev. 87:217-221(1999).

CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.

```

DR EMBL: AJ237750; CAC20109.1;
DR HSSP: P01308; ILPH;
DR ZFIN: ZDB-GENE-980526-110; Ins.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR00277; INSULINB.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
KW Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 53 INSULIN B CHAIN.
FT CHAIN 86 108 INSULIN A CHAIN.
SQ SEQUENCE 108 AA; 11904 MW; 3195289E72AD6D25 CRC64;

Query Match 33.3%; Score 195.5; DB 13; Length 108;
Best Local Similarity 45.1%; Pred. No. 2e-15;
Matches 37; Conservative 5; Mismatches 7; Indels 33; Gaps 1;

QY 59 QHLCGSHLYEALVLCGERGFYTPK-----T 85
DQ 27 QHLCGSHLYEALVLCGPTGFYFNPKROVEPLLGFLPKSAQETEVADFATKDHAEILRK 86

QY 86 RGIVEQCCISGLYLENYCN 107
DQ 87 RGIVEQCCCHKPCISFELQNYCN 108

RESULT 13
Q902N4 PRELIMINARY; PRT: 108 AA.
AC Q902N4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Preproinsulin.
OS Catla catla (Catla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Catla.
OX NCBI_TaxID=72446;
RN [1]
RP SEQUENCE FROM N.A.
RA Bhattacharya S., Roy S.S., Dasgupta S., Ravikumar L., Mukherjee M.,
RA Bandyopadhyaya I., Wakabayashi K.;
RT "A new cell secreting insulin.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL: AF373021; AAK5158.1;
DR HSSP: P01308; ILNP.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR00277; INSULINB.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
SQ SEQUENCE 108 AA; 11881 MW; D713036E22EF5D59 CRC64;

Query Match 33.3%; Score 195.5; DB 13; Length 108;
Best Local Similarity 45.1%; Pred. No. 2e-15;
Matches 37; Conservative 5; Mismatches 7; Indels 33; Gaps 1;

QY 59 QHLCGSHLYEALVLCGERGFYTPK-----T 85
DQ 27 QHLCGSHLYEALVLCGPTGFYFNPKROVEPLLGFLPKSAQETEVADFATKDHAEILRK 86

QY 86 RGIVEQCCISGLYLENYCN 107
DQ 87 RGIVEQCCCHKPCISFELQNYCN 108

RESULT 14
Q98TB0 PRELIMINARY; PRT: 111 AA.
ID Q98TB0
QY 2 FTPIPLSRFDNAHLRAHLHQLAFDTYQEFEPAYIPKQKYSFLQNP 49

```

```

AC Q98TB0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Preproinsulin (Fragment).
OS Chitala chitala (Clown knifefish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;
OC Osteoglossiformes; Notopteridae; Chitala.
OX NCBI_TaxID=112153;
RN [1]
RP SEQUENCE FROM N.A.
RA Al-Mahrouki A.A., Irwin D.M., Ibrahim L.C., Youson J.H.;
RT "Molecular cloning of preproinsulin cDNAs from several
RT osteoglossomorphs and a cyprinid.";
RL Mol. Cell. Endocrinol. 174:58 (2001).
CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL: AF199586; AAK28710.1;
DR HSSP: P01308; ILPH.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR SMART: SM00078; IIGF; 1.
DR NON_TER 111
FT NON_TER 111
SQ SEQUENCE 111 AA; 12483 MW; 247CA4431376329F CRC64;

Query Match 33.2%; Score 195; DB 13; Length 111;
Best Local Similarity 44.2%; Pred. No. 2.4e-15;
Matches 38; Conservative 3; Mismatches 9; Indels 46; Gaps 1;

QY 58 NQHLGSHLYEALVLCGERGFYTPK-----84
DQ 26 NQHLGSHLYEALVLCGERGFYFNPKMDKRAEPLLGFLSPKSGLENEVDYEPFKDQKD 85

QY 85 ---TRGIVEQCCISGLYLENYCN 107
DQ 86 VKMKRGIVEQCCHRPCNIFQDQNYCN 111

RESULT 15
Q8WND9 PRELIMINARY; PRT: 217 AA.
ID Q8WND9
AC Q8WND9;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Growth hormone.
GN GH-V.
OS Ateles geoffroyi (Black-handed spider monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Ateles.
OX NCBI_TaxID=9509;
RN [1]
RP SEQUENCE FROM N.A.
RA Revol A., Esquivel D., Santiago D., Barrera-Saldana H.
RT "Independent duplication of the growth hormone gene in three
RT Anthropoidcan lineages.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF374235; AAL72287.1;
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone; 1.
DR PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
SQ SEQUENCE 217 AA; 25293 MW; 741745A1B75C053E CRC64;

Query Match 33.2%; Score 195; DB 6; Length 217;
Best Local Similarity 77.1%; Pred. No. 5.2e-15;
Matches 37; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 2 FTPIPLSRFDNAHLRAHLHQLAFDTYQEFEPAYIPKQKYSFLQNP 49

```

Ob 27 FPRIPLSRFGDAMLRAHQLHQAFTYQLELENCIPKKQKYFFLRNP 74

Search completed: September 16, 2003, 12:40:00  
Job time : 27.8132 secs

GenCore version 1.6  
Copyright (c) 1993 - 2003 (mpugen Ltd.)

OM protein - protein search, using sw model

Run on: September 16, 2003, 12:11:20 ; Search time 44.358 Seconds  
(without alignments)  
536.746 Million cell updates/sec

Title: US-09-423-100-7

Perfect score: 797

Sequence: 1 MFPTPLRLFDNMLRAHR.....IVQCCTCTCSLYQLNLYCN 150

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03:\*

```

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	797	100.0	150	20 AAY42861	Chimeric protein,
2	555.5	69.7	107	20 AAY42860	hGH mini-proinsulin
3	470	59.0	92	20 AAY42856	Human growth hormo
4	470	59.0	134	20 AAY52285	Human anti-angioge
5	470	59.0	191	23 ABG94861	Human growth hormo
6	466	58.5	192	10 AAP90129	Human growth hormo
7	466	58.5	192	20 AAW92264	Human anti-angioge
8	465	58.3	140	10 AAP91041	Human growth hormo
9	465	58.3	261	10 AAP91299	Human nerve growth

10	465	58.3	262	12 AAR11740	Human growth hormo
11	465	58.3	310	11 AAR03255	Fusion protein of
12	464	58.2	191	23 ABG31862	Mature human growt
13	463	58.1	191	23 ABG94860	Human growth hormo
14	463	58.1	191	23 ABG94977	Human growth hormo
15	462	58.0	144	11 AAR05313	Segment of B-cell
16	462	58.0	191	23 ABG94975	Human growth hormo
17	462	58.0	191	23 ABG94976	Human growth hormo
18	462	58.0	262	7 AAP51033	Human growth hormo
19	461	57.8	191	18 AAO20110	Protein sequence o
20	461	57.8	191	20 AAY04396	Natural human 22kD
21	461	57.8	191	21 AAY78425	Human growth hormo
22	461	57.8	191	22 AAO17485	Human growth hormo
23	461	57.8	191	22 AAO17486	Human growth hormo
24	461	57.8	191	23 ABG94853	Human growth hormo
25	461	57.8	191	23 ABG94864	Human growth hormo
26	461	57.8	191	23 ABG94907	Human growth hormo
27	461	57.8	191	23 ABG94908	Human growth hormo
28	461	57.8	191	23 ABG94909	Human growth hormo
29	461	57.8	191	24 ABG94910	Human growth hormo
30	461	57.8	191	23 ABG94911	Human growth hormo
31	461	57.8	191	23 ABG94912	Human growth hormo
32	461	57.8	191	23 ABG94913	Human growth hormo
33	461	57.8	191	23 ABG94914	Human growth hormo
34	461	57.8	191	23 ABG94915	Human growth hormo
35	461	57.8	191	23 ABG94916	Human growth hormo
36	461	57.8	191	23 ABG94917	Human growth hormo
37	461	57.8	191	23 ABG94918	Human growth hormo
38	461	57.8	191	23 ABG94919	Human growth hormo
39	461	57.8	191	23 ABG94920	Human growth hormo
40	461	57.8	191	23 ABG94921	Human growth hormo
41	461	57.8	191	23 ABG94922	Human growth hormo
42	461	57.8	191	23 ABG94923	Human growth hormo
43	461	57.8	191	23 ABG94924	Human growth hormo
44	461	57.8	191	23 ABG94925	Human growth hormo
45	461	57.8	191	23 ABG94926	Human growth hormo

#### ALIGNMENTS

```

RESULT 1
ID AAY42861 standard; protein: 150 AA.
XX AC AAY42861;
XX 19-JAN-2000 (first entry)
XX DE Chimeric protein, SEQ ID 1.
XX KW Insulin; precursor; growth hormone; chaperone; intramolecular;
XX folding; conformation; chimeric protein; cleavable; recombinant;
XX production; yield.
XX OS Synthetic.
XX OS Homo sapiens.
XX PN WO9950302-A1.
XX PD 07-OCT-1999.
XX PF 31-MAR-1998; 98WO-CN00052.
XX PR 31-MAR-1998; 98WO-CN00052.
XX PA (TONG-) TONGHUA GANTECH BIOTECHNOLOGY LTD.
XX PI Can Z;
XX DR WPI; 1999-610839/52.
XX PT New chimeric proteins containing human growth hormone fragment, used

```

PT particularly for the production of human insulin  
 XX  
 PS Claim 14; Page 30-31; 46pp; English.

CC This sequence represents a chimeric protein, which contains an  
 CC N-terminal fragment of human growth hormone (hGH) of the sequence given  
 CC in AAY42856, a cleavable peptide linker (AAY42857), and a human insulin  
 CC precursor comprising insulin A and B chains (AAY42859). The hGH portion  
 CC of the chimeric protein acts as an intramolecular chaperone (IMC) for  
 CC the insulin precursor, enabling it to fold correctly. The cleavable  
 CC peptide linker has a C-terminal Arg residue which enables the hGH portion  
 CC of the chimeric protein to be removed after folding has taken place.  
 CC Production of recombinant human insulin via an hGH-proinsulin chimeric  
 CC protein can provide human insulin with correctly linked cysteine bridges  
 CC with fewer necessary procedural steps, and hence resulting in a higher  
 CC yield of human insulin. The IMC sequences not only protect insulin  
 CC sequences from intracellular degradation by a microorganism host, but  
 CC also promote the folding of the fused insulin precursor, facilitate the  
 CC solubility of the fusion protein and decrease the intermolecular  
 CC interactions among the fusion proteins, thus allowing folding of the  
 CC fused insulin precursor at commercially useful high concentrations. The  
 CC procedural steps of cyanogen bromide cleavage, oxidative sulphytolysis  
 CC and related purification steps can thus be eliminated, along with the use  
 CC of high concentrations of mercaptan or the use of hydrophobic absorbent  
 CC resins.

XX Sequence 150 AA;

Query Match 100.0%; Score 797; DB 20; Length 150;  
 Best Local Similarity 100.0%; Pred. No. 1e-42;  
 Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFPTPLSRFLDNAMLRHLRLHQLAFDTYQEFEEAYIPKEQKYSFLONPQTSLSFSESIP 60  
 DB 1 MFPTPLSRFLDNAMLRHLRLHQLAFDTYQEFEEAYIPKEQKYSFLONPQTSLSFSESIP 60  
 QY 61 TPSNREETQOKSNLELLRISLLLSQSLWLEPVOLGTGPRFVNOHLCGSHLYEALVLCGER 120  
 DB 61 TPSNREETQOKSNLELLRISLLLSQSLWLEPVOLGTGPRFVNOHLCGSHLYEALVLCGER 120  
 QY 121 GFFYTPKTRGIVEQCCTSLCYOLENYCN 150  
 DB 121 GFFYTPKTRGIVEQCCTSLCYOLENYCN 150

# RESULT 2

AA42860  
 ID AAY42860 standard; protein: 107 AA.

XX AAY42860;

DT 19-JAN-2000 (first entry)

XX hGH-mini-proinsulin chimeric protein.

DE Insulin; precursor; growth hormone; chaperone; intramolecular;  
 KW folding; conformation; chimeric protein; cleavable; recombinant;  
 KW production; yield.

XX Synthetic.

OS Homo sapiens.

XX W09950302-A1.

PN 07-OCT-1999.

XX 31-MAR-1998; 98WO-CN00052.

XX 31-MAR-1998; 98WO-CN00052.

XX (TONG-) TONGHUA GANTECH BIOTECHNOLOGY LTD.

PA Gan 2;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX WPI; 1999-610839/52.

XX New chimeric proteins containing human growth hormone fragment, used  
 PT particularly for the production of human insulin

XX Claim 13; Page 30; 46pp; English.

XX This sequence represents a chimeric protein, hGH-mini-proinsulin.  
 CC This chimeric protein contains an N-terminal fragment of human growth  
 CC hormone (hGH) of the sequence given in AAY42855, a cleavable peptide  
 CC linker (AAY42857), and a human insulin precursor comprising insulin  
 CC A and B chains (AAY42859). The hGH portion of the chimeric protein acts  
 CC as an intramolecular chaperone (IMC) for the insulin precursor,  
 CC enabling it to fold correctly. The cleavable peptide linker has a  
 CC C-terminal Arg residue which enables the hGH portion of the  
 CC chimeric protein to be removed after folding has taken place. Production  
 CC of recombinant human insulin via an hGH-proinsulin chimeric protein can  
 CC provide human insulin with correctly linked cysteine bridges with  
 CC fewer necessary procedural steps, and hence resulting in a higher yield  
 CC of human insulin. The IMC sequences not only protect insulin sequences  
 CC from intracellular degradation by a microorganism host, but also promote  
 CC the folding of the fused insulin precursor, facilitate the solubility of  
 CC the fusion protein and decrease the intermolecular interactions among  
 CC the fusion proteins, thus allowing folding of the fused insulin precursor  
 CC at commercially useful high concentrations. The procedural steps of  
 CC cyanogen bromide cleavage, oxidative sulphytolysis and related  
 CC purification steps can thus be eliminated, along with the use of high  
 CC concentrations of mercaptan or the use of hydrophobic absorbent resins.

XX Sequence 107 AA;

Query Match 69.7%; Score 555.5; DB 20; Length 107;  
 Best Local Similarity 71.3%; Pred. No. 8e-28;  
 Matches 107; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MFPTPLSRFLDNAMLRHLRLHQLAFDTYQEFEEAYIPKEQKYSFLONPQTSLSFSESIP 60  
 DB 1 MFPTPLSRFLDNAMLRHLRLHQLAFDTYQEFEEAYIPKEQKYSFLONPQTSLSFSESIP 60  
 QY 61 TPSNREETQOKSNLELLRISLLLSQSLWLEPVOLGTGPRFVNOHLCGSHLYEALVLCGER 120  
 DB 50 -----LGTGPRFVNOHLCGSHLYEALVLCGER 77

QY 121 GFFYTPKTRGIVEQCCTSLCYOLENYCN 150

DB 78 GFFYTPKTRGIVEQCCTSLCYOLENYCN 107

# RESULT 3

AA42856

ID AAY42856 standard; protein: 92 AA.

XX AAY42856;

DT 19-JAN-2000 (first entry)

DE Human growth hormone (hGH) N-terminal fragment #2.

KW Growth hormone; chaperone; intramolecular; insulin; precursor;  
 KW folding; conformation; chimeric protein; cleavable; recombinant;  
 KW production; yield.

XX Homo sapiens.

OS W09950302-A1.

PN 07-OCT-1999.

XX 31-MAR-1998; 98WO-CN00052.

XX 31-MAR-1998; 98WO-CN00052.

XX

XX

XX

XX

XX

XX

XX

XX

PA (TONG-) TONGCHUA GANTECH BIOTECHNOLOGY LTD.

PI Gan 2;

XX WPI; 1999-610839/52.

XX New chimeric proteins containing human growth hormone fragment, used particularly for the production of human insulin

XX Claim 5; Page 28; 46pp; English.

XX This sequence represents an N-terminal fragment of human growth hormone (hGH) which is a component of a chimeric protein (AA42861) which also contains a human insulin precursor (AA42859). The hGH portion of the chimeric protein acts as an intramolecular chaperone (IMC) for the insulin precursor, enabling it to fold correctly. A cleavable peptide linker with a C-terminal Arg residue (AA42857) enables the hGH portion of the chimeric protein to be removed after folding has taken place. CC Production of recombinant human insulin via an hGH-proinsulin chimeric protein can provide human insulin with correctly linked cysteine bridges CC with fewer necessary procedural steps, and hence resulting in a higher CC yield of human insulin. The IMC sequences not only protect insulin CC sequences from intracellular degradation by a microorganism host, but CC also promote the folding of the fused insulin precursor, facilitate the CC solubility of the fusion protein and decrease the intermolecular CC interactions among the fusion proteins, thus allowing folding of the CC fused insulin precursor at commercially useful high concentrations. The CC procedural steps of cyanogen bromide cleavage, oxidative sulphytolysis CC and related purification steps can thus be eliminated, along with the CC use of high concentrations of mercaptan or the use of hydrophobic CC absorbent resins.

XX SQ Sequence 92 AA;

Query Match 59.0%; Score 470; DB 20; Length 92;

Best Local Similarity 100.0%; Pred. No. 1.5e-22;

Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFPTPLSLRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPTLSLFSSESIP 60

Db 1 MFPTPLSLRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPTLSLFSSESIP 60

QY 61 TPSNREETOQKSNLELLRISLLLIQSWLEPVQ 92

Db 61 TPSNREETOQKSNLELLRISLLLIQSWLEPVQ 92

RESULT 4

AAW92265

XX AAW92265 standard; Protein: 134 AA.

XX AC AAW92265;

XX DT 08-JUN-1999 (first entry)

XX Human anti-angiogenic peptide 16K hGH Met-1pro133.

XX Human; anti-angiogenic; prolactin; placental lactogen; hPL; angiogenesis; growth hormone; hGH; hGH-V; capillary endothelial cell proliferation; placental vascularisation; pregnancy; treatment; angiogenic disease; tumour; inhibitor; malignant; angiofibroma; arteriovenous malformation; arthritis; atherosclerotic plaques; corneal graft neovascularisation; wound healing; proliferative retinopathy; macular degeneration; trachoma; psoriasis; glaucoma; ocular; uveitis; fracture; Osler-Weber syndrome; ulcer; leukaemia; reproductive disorder; Kaposi's sarcoma; vascular adhesion; gene therapy; pre-eclampsia; intrauterine growth retardation; placental dysfunction.

XX Homo sapiens.

XX OS W09851323-A1.

XX PN

PD 19-NOV-1998.

PF 12-MAY-1998; 98WO-US09691.

XX 13-MAY-1997; 97US-0046394.

PR (REGC ) UNIV CALIFORNIA.

XX Martial JA, Struman I, Taylor R, Weiner RI;

XX WPI; 1999-045192/04.

XX N-PSDB; AAX01707.

XX New anti-angiogenic peptides - comprise N-terminal fragments of human placental lactogen, human growth hormone, growth hormone variant or human prolactin

XX Claim 4; Page 49-50; 87pp; English.

XX This invention describes novel human anti-angiogenic peptides derived from 10 to 150 consecutive amino acids selected from the N-terminal end of human placental lactogen (hPL), human growth hormone (hGH), growth hormone variant (hGH-V), or human prolactin. Such peptides (i) inhibit capillary endothelial cell proliferation and organisation (ii) inhibit angiogenesis in chick chorioallantoic membrane and (iii) binds to at least one specific receptor which does not bind an intact full length hGH, hPL, prolactin or hGH-V. The invention also describes a method for diagnosing a probable abnormality of placental vascularisation during pregnancy. The peptides can be used for treating an angiogenic disease in a subject, for inhibiting tumour formation or growth in a patient or for modulating vascularisation of a patient's placenta. In particular, the peptides can be used for preventing or treating e.g. malignant tumours, angiofibroma, arteriovenous malformation, arthritic such as rheumatoid arthritis, atherosclerotic plaques, corneal graft neovascularisation, delayed wound healing, proliferative retinopathy such as diabetic retinopathy, macular degeneration, granulations such as those occurring in haemophilic joints, inappropriate vascularisation in wound healing such as hypertrophic scars or keloid scars, neovascular glaucoma, ocular tumour, uveitis, non-union fractures, Osler-Weber syndrome, psoriasis, pyogenic glaucoma, retrolental fibroplasia, scleroderma, solid tumours, Kaposi's sarcoma, trachoma, vascular adhesions, chronic varicose ulcers, leukaemia, and reproductive disorders such as follicular and luteal cysts and choriocarcinoma. They can also be used as contraceptive agents. DNA encoding the peptides can be used in gene therapy. The measurement of abnormal levels of N-terminal fragments of hGH, hGH-V, prolactin or hPL can be used in assays for impairment of vascular development associated with pre-eclampsia, intrauterine growth retardation, and placental dysfunction.

XX SQ Sequence 134 AA;

Query Match 59.0%; Score 470; DB 20; Length 134;

Best Local Similarity 100.0%; Pred. No. 2e-22;

Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFPTPLSLRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPTLSLFSSESIP 60

Db 1 MFPTPLSLRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPTLSLFSSESIP 60

QY 61 TPSNREETOQKSNLELLRISLLLIQSWLEPVQ 92

Db 61 TPSNREETOQKSNLELLRISLLLIQSWLEPVQ 92

RESULT 5

ABG94861

ID ABG94861 standard; Protein: 191 AA.

XX AC ABG94861;

XX DT 03-DEC-2002 (first entry)

XX Human growth hormone mutant hPRL (111-129).

DE

XX Growth hormone; placental lactogen; prolactin; active domain; hGH;  
 KW structure-function relationship; segment-substituted polypeptide;  
 KW mutant; mutein.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 PN US6428954-B1.  
 PD 06-AUG-2002.  
 XX  
 PF 06-JUN-1995; 95US-0483039.  
 XX  
 PR 26-OCT-1989; 89US-0428066.  
 PR 27-APR-1992; 92US-0875204.  
 PR 13-APR-1992; 92US-0960227.  
 PR 02-FEB-1994; 94US-0190723.  
 PR 28-OCT-1988; 88US-0264611.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX  
 PI Wells JA, Cunningham BC;  
 XX  
 DR WPI: 2002-696875/75.  
 XX  
 PT Identifying active domains within cloned polypeptides of known amino  
 PT acid sequence by substituting analog segments into the parent  
 PT polypeptide is useful to determine the relationship between structure  
 PT and function.  
 XX  
 PS Example 1; Page -: 86pp; English.  
 XX  
 CC The invention relates to identifying an unknown active domain in a region  
 CC of known amino acid sequence in a parent polypeptide e.g. human growth  
 CC hormone (hGH) which has been cloned and has a pre-identified biological  
 CC activity, where the active domain interacts with a target when the parent  
 CC polypeptide is in its native-folded form and the interaction is  
 CC responsible for the biological activity comprising: (a) comparing the  
 CC amino acid sequence or polypeptide structure in the region of known amino  
 CC acid sequence of hGH with the amino acid sequence or polypeptide  
 CC structure in a region of known amino acid sequence of an analogue  
 CC polypeptide (e.g. prolactin, placental lactogen or porcine growth  
 CC hormone) which has at least 15% homology with hGH alpha-carbon  
 CC coordinates within about 2-3.5 angstroms of hGH alpha-carbon  
 CC for about 60% of the analogue sequence, where any interaction of the  
 CC analogue with the target is different from target interaction with hGH;  
 CC (b) substituting DNA encoding an analogous polypeptide segment from the  
 CC analogue into DNA encoding the full length hGH, and expressing a  
 CC segment-substituted polypeptide; (c) contacting the segment-substituted  
 CC polypeptide with the target to determine interaction; (d) repeating steps  
 CC (b) and (c) with a second analogous polypeptide segment; and  
 CC (e) comparing the difference between activity of the first and second  
 CC segment-substituted polypeptides as an indication of the location of  
 CC the unknown active domain in hGH. The method is useful for determining  
 CC the relationship between structure and function of known polypeptide  
 CC sequences. The present sequence is that of human growth hormone  
 CC mutant substituted with residues from an hGH analogue (prolactin,  
 CC placental lactogen or porcine growth hormone).  
 CC Note: The present sequence is not shown in the specification but was  
 CC created by the indexer using the mature hGH sequence and information  
 CC contained in the specification.  
 XX  
 SQ Sequence 191 AA;

Query Match 59.0%; Score 470; DB 23; Length 191;  
 Best Local Similarity 69.2%; Pred. No. 2.6e-22;  
 Matches 101; Conservative 8; Mismatches 19; Indels 18; Gaps 3;  
 QY 2 FPTPLSLFDNMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPQTSLSFSSEIPT 61  
 DB 1 FPTPLSLFDNMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPQTSLSFSSEIPT 60

QY 62 PSNREETOQKSNLELLRLISLLLIQSWLEPVLGTGPRFVNOHLCGS-----H 108  
 DB 61 PSNREETOQKSNLELLRLISLLLIQSWLEPVLGTGPRFVNOHLCGS-----H 119  
 QY 109 LVEALYLVCGERGFFYTPKTRGIVEQ 134  
 DB 120 LIEGLMLILSDG-----SPRT; IFKQ 141  
 RESULT 6  
 AAP90129 standard: protein; 192 AA.  
 ID XX  
 AC AAP90129;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 06-FEB-1996 (revised)  
 DT 01-NOV-1989 (first entry)  
 XX  
 DE Human growth hormone.  
 XX  
 KW Human growth hormone; fusion protein; recombinant  
 KW vector.  
 KW  
 OS Homo sapiens (Human).  
 XX  
 PN JP01144981-A.  
 XX  
 PD 07-JUN-1989.  
 XX  
 PF 02-DEC-1987; 87JP-0304937.  
 XX  
 PR 02-DEC-1987; 87JP-0304937.  
 XX  
 PA (WAKI ) WAKUNAGA SEIYAKU KK.  
 XX  
 DR WPI: 1989-209284/29.  
 DR N-PSDB; AAN90269.  
 XX  
 PT Recombinant vector contg. fus. protein - consisting of human  
 PT growth hormone or deriv. lig. 1 to foreign protein, for stability  
 PT and high yield.  
 XX  
 PS Disclosure: Fig 1; 19pp; Japanese.  
 XX  
 CC The invention consists of a vector contg. a fusion protein which is  
 CC formed by ligating, downstream of a promoter, hGH or a deriv. (pref.  
 CC formed by substn. of Met-14 with leu) and a foreign protein.  
 CC Stability of the vector in the host is greatly increased so the  
 CC protein yield is higher.  
 CC (Updated on 25-MAR-2003 to correct PA field.)  
 XX  
 SQ Sequence 192 AA;  
 Query Match 58.5%; Score 466; DB 10; Length 192;  
 Best Local Similarity 70.5%; Pred. No. 4.6e-22;  
 Matches 103; Conservative 7; Mismatches 20; Indels 16; Gaps 4;  
 QY 1 MFPTPLSLFDNMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPQTSLSFSSEIP 60  
 DB 1 MFPTPLSLFDNMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPQTSLSFSSEIP 60  
 QY 61 TPSNREETOQKSNLELLRLISLLLIQSWLEPVLGTGPRFVNOHLCGS-----H 110  
 DB 61 TPSNREETOQKSNLELLRLISLLLIQSWLEPVLGTGPRFVNOHLCGS-----H 119  
 QY 111 EALYLVCGERGFFYTPKTRGIVEQ 134  
 DB 120 EGIOTLMGRLEDC---SPRTGQIFKQ 142  
 RESULT 7  
 AAN92264

ID AAW92264 standard; Protein: 192 AA.

XX AC AAW92264;

XX DT 08-JUN-1999 (first entry)

XX DE Human anti-angiogenic peptide hGH Met-1pHe191.

XX KW Human; anti-angiogenic; prolactin; placental lactogen; hPL; angiogenesis;  
 KW growth hormone; hGH; hGH-V; capillary endothelial cell proliferation;  
 KW placental vasculature; pregnancy; treatment; angiogenic disease;  
 KW tumour; inhibitor; malignant; angiofibroma; arteriovenous malformation;  
 KW arthritis; atherosclerotic plaques; corneal graft neovascularisation;  
 KW wound healing; proliferative retinopathy; macular degeneration; trachoma;  
 KW granulation; glaucoma; ocular; uveitis; fracture; Osler-Weber syndrome;  
 KW psoriasis; fibroplasia; scleroderma; Kaposi's sarcoma; vascular adhesion;  
 KW ulcer; leukaemia; reproductive disorder; contraceptive agent;  
 KW gene therapy; pre-eclampsia; intrauterine growth retardation;  
 KW placental dysfunction.

XX OS Homo sapiens.

XX PN W09851323-Al.

XX PD 19-NOV-1998.

XX PF 12-MAY-1998; 98MO-US09691.

XX PR 13-MAY-1997; 97US-0046394.

XX PA (REGC ) UNIV CALIFORNIA.

XX PI Martial JA, Struman I, Taylor R, Weiner RI;

XX DR WPI: 1999-045192/04.

XX DR N-PSDB; AAX01706.

XX PT New anti-angiogenic peptides - comprise N-terminal fragments of  
 PT human placental lactogen, human growth hormone, growth hormone  
 PT variant or human prolactin

XX PS Example 3; Page 49; 87pp; English.

XX CC This invention describes novel human anti-angiogenic peptides derived  
 CC from 10 to 150 consecutive amino acids selected from the N-terminal end  
 CC of human placental lactogen (hPL), human growth hormone (hGH), growth  
 CC hormone variant (hGH-V), or human prolactin. Such peptides (i) inhibit  
 CC capillary endothelial cell proliferation and organisation (ii) inhibit  
 CC angiogenesis in chick chorioallantoic membrane and (iii) binds to at  
 CC least one specific receptor which does not bind an intact full length  
 CC hGH, hPL, prolactin or hGH-V. The invention also describes a method for  
 CC diagnosing a probable abnormality of placental vascularisation during  
 CC pregnancy. The peptides can be used for treating an angiogenic disease in  
 CC a subject, for inhibiting tumour formation or growth in a patient or for  
 CC modulating vascularisation of a patient's placenta. In particular, the  
 CC peptides can be used for preventing or treating e.g. malignant tumours,  
 CC angiofibroma, arteriovenous malformation, arthritic such as rheumatoid  
 CC arthritis, atherosclerotic plaques, corneal graft neovascularisation,  
 CC delayed wound healing, proliferative retinopathy such as diabetic  
 CC retinopathy, macular degeneration, granulations such as those occurring  
 CC in haemophilic joints, inappropriate vascularisation in wound healing  
 CC such as hypertrophic scars or keloid scars, neovascular glaucoma, ocular  
 CC tumour, uveitis, non-union fractures, Osler-Weber syndrome, psoriasis,  
 CC pyogenic glaucoma, retrolental fibroplasia, scleroderma, solid tumours,  
 CC Kaposi's sarcoma, trachoma, vascular adhesions, chronic varicose ulcers,  
 CC leukaemia, and reproductive disorders such as follicular and luteal cysts  
 CC and choriocarcinoma. They can also be used as contraceptive agents. DNA  
 CC encoding the peptides can be used in gene therapy. The measurement of  
 CC abnormal levels of N-terminal fragments of hGH, hGH-V, prolactin or hPL  
 CC can be used in assays for impairment of vascular development associated  
 CC with pre-eclampsia, intrauterine growth retardation, and placental  
 CC dysfunction.

XX

SQ Sequence 192 AA;

Query Match 58.58; Score 466; DB 20; Length 192;

Best Local Similarity 70.58; Pred. No. 4.2e-22;

Matches 103; Conservative 7; Mismatches 20; Indels 16; Gaps 4;

Oy 1 MFPTIPLSRLFDNAMLRAH; HQLAFTDYQEFEEAYIPKEQKYSFLONPOTLSLFSSESIP 60

Db 1 MFPTIPLSRLFDNAMLRAH; HQLAFTDYQEFEEAYIPKEQKYSFLONPOTLSLFSSESIP 60

Oy 61 TPSNRETOQKSNLELLRISLLLIQSWLEPVQIGTGRFFVNOHLGGS-----HLV 110

Db 61 TPSNRETOQKSNLELLRISLLLIQSWLEPVQF-LRSVFANSLVYGASISNVYDLKDLF 119

Oy 111 EALYLVGS--ERGFFYTPKTRGIVEQ 134

Db 120 EGIOITLMGRLEDG---SPRTQOIFKQ 142

RESULT 8

AAP91041

ID AAP91041 standard; protein: 40 AA.

XX AC AAP91041;

XX DT 14-DEC-1989 (first entry).

XX DE Human growth hormone segment.

XX KW Human growth hormone; fusion protein; thrombin;

XX KW geriatric dementia; nervous disorders; human nerve factor.

XX OS Homo sapiens (human).

XX PN EP329175-A.

XX PD 23-AUG-1989.

XX PF 17-FEB-1989; 89EP-0102795.

XX PR 19-FEB-1988; 88JP-0035042.

XX PA (TOYJ ) TOSOH CORP.

XX PI Ohtsuka E.

XX WPI: 1989-243092/34.

XX PT New human nerve growth factor gene encoding fusion protein

XX PT - having cleavage site for thrombin, useful for treating geriatric

XX PT dementia, etc.

XX PS Disclosure: page 21; 38pp; English.

XX CC Human growth hormone segment, used at the N-terminal of a fusion

XX CC protein, which contains a thrombin recognition site, and human beta nerve

XX CC growth factor (beta-NGF) at the C-terminal. Beta-NGF can be used to

XX CC control geriatric dementia and other nervous disorders, and can be

XX CC released from the fusion protein by incubation with thrombin (see

XX CC AAN90577-8, AAP91034, AAP91299).

XX SQ Sequence 140 AA;

Query Match 58.38; Score 465; DB 10; Length 140;

Best Local Similarity 98.98; Pred. No. 4.2e-22;

Matches 91; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 MFPTIPLSRLFDNAMLRAH; HQLAFTDYQEFEEAYIPKEQKYSFLONPOTLSLFSSESIP 60

Db 1 MFPTIPLSRLFDNAMLRAH; HQLAFTDYQEFEEAYIPKEQKYSFLONPOTLSLFSSESIP 60

Oy 61 TPSNRETOQKSNLELLRISLLLIQSWLEPVQ 92

Db 61 TPSNREETOOKSNLELLRISLLLIQSWLEPVQ 92

## RESULT 9

AAP91299  
ID AAP91299 standard; protein; 261 AA.

XX AC AAP91299;

XX DT 14-DEC-1989 (first entry).

XX DE Human nerve growth factor and human growth hormone fusion protein.

XX KW Human nerve growth factor; fusion protein; thrombin;

XX KW geriatric dementia; nervous disorders; human growth hormone.

XX OS Homo sapiens (human).

XX FH Key Location/Qualifiers

FT Region 1..140

FT Region 141..143

FT Region 144..261

XX PN EP329175-A.

XX PD 23-AUG-1989.

XX PF 17-FEB-1989; 89EP-0102795.

XX PR 19-FEB-1988; 88JP-0035042.

XX PA (TOYJ ) TOSOH CORP.

XX PI Ohtsuka E;

XX DR WPI; 1989-243092/34.

XX PT New human nerve growth factor gene encoding fusion protein

XX PT - having cleavage site for thrombin, useful for treating geriatric

XX PS Claim 36; page 31-32; 38pp; English.

XX CC Fusion protein consisting of human growth hormone at the  
N-terminal end (1st region); a 3 amino acid sequence representing  
thrombin recognition site; and human beta nerve growth factor (beta-NGF)  
at the C-terminal. Beta-NGF can be used to control geriatric dementia  
and other nervous disorders, and can be released from the fusion  
protein by incubation with thrombin (see AAN90577-8, AAP91034).

XX SQ Sequence 261 AA;

Query Match 58.3%; Score 465; DB 10; Length 261;

Best Local Similarity 98.9%; Pred. No. 6.8e-22;

Matches 91; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MFPTIPLSLRFDNAMLRAHRLHQLAFDTYOEFEEAYIPKEQKYSFLQNPTSLSFSESIP 60

DB 1 MFPTIPLSLRFDNAMLRAHRLHQLAFDTYOEFEEAYIPKEQKYSFLQNPTSLSFSESIP 60

QY 61 TPSNREETOOKSNLELLRISLLLIQSWLEPVQ 92

DB 61 TPSNREETOOKSNLELLRISLLLIQSWLEPVQ 92

## RESULT 10

AAR11740

ID AAR11740 standard; Protein; 262 AA.

XX AC AAR11740;

XX DT 25-MAR-2003 (updated)

DT 25-JUN-1991 (first entry)

XX DE Human growth hormone/human nerve growth factor beta fusion protein.

XX KW hGH; hNGF; nervous system diseases; dementia.

XX OS Homo sapiens.

XX PN JP03067598-A.

XX PD 22-MAR-1991.

XX PF 07-AUG-1989; 89JP-0202835.

XX PR 07-AUG-1989; 89JP-0202835.

XX PA (TOYJ ) TOSOH CORP.

XX DR WPI; 1991-128768/18.

XX DR N-PSDB; AAQ11578.

XX PT Purification of human neuron growth factor beta subunit-contg. protein -  
by contacting with gel having cation exchange gp. in presence of  
urea

XX PS Disclosure ; fig 1; 7pp; Japanese.

XX CC A recombinant human nerve growth factor beta subunit-contg. protein  
can be produced as this fusion protein. It is purified by contacting  
a gel having a cation exchange gp. with the fusion protein, in the  
presence of urea. The purified protein is useful in a medicament  
for treating disorders of the nervous system, eg dementia.  
(Updated on 25-MAR-2003 to correct PF field.)

XX SQ Sequence 262 AA;

Query Match 58.3%; Score 465; DB 12; Length 262;

Best Local Similarity 98.9%; Pred. No. 6.8e-22;

Matches 91; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MFPTIPLSLRFDNAMLRAHRLHQLAFDTYOEFEEAYIPKEQKYSFLQNPTSLSFSESIP 60

DB 1 MFPTIPLSLRFDNAMLRAHRLHQLAFDTYOEFEEAYIPKEQKYSFLQNPTSLSFSESIP 60

QY 61 TPSNREETOOKSNLELLRISLLLIQSWLEPVQ 92

DB 61 TPSNREETOOKSNLELLRISLLLIQSWLEPVQ 92

## RESULT 11

AAR03255

ID AAR03255 standard; protein; 310 AA.

XX AC AAR03255;

XX DT 19-JUL-1990 (first entry)

XX DE Fusion protein of B-cell stimulatory factor-2 and B-cell

XX DE differentiation factor.

XX KW B-cell stimulatory factor-2; interleukin-6; B-cell differentiation;

XX KW interleukin-5; fusion protein.

XX OS Homo sapiens.

XX PN JP02013375-A.

XX PD 17-JAN-1990.

XX PF 01-JUL-1988; 88JP-0162556.

XX PR 01-JUL-1988; 88JP-0162556.



DR WPI: 2002-696875/75.  
 XX  
 PT Identifying active domains within cloned polypeptides of known amino  
 PT acid sequence by substituting analog segments into the parent  
 PT polypeptide is useful to determine the relationship between structure  
 PT and function.  
 XX  
 PS Example 1; Page -: 86pp; English.  
 PS  
 XX The invention relates to identifying an unknown active domain in a region  
 CC of known amino acid sequence in a parent polypeptide e.g. human growth  
 CC hormone (hGH) which has been cloned and has a pre-identified biological  
 CC activity, where the active domain interacts with a target when the parent  
 CC polypeptide is in its native-folded form and the interaction is  
 CC responsible for the biological activity comprising: (a) comparing the  
 CC amino acid sequence or polypeptide structure in the region of known amino  
 CC acid sequence of hGH with the amino acid sequence of an analogue  
 CC polypeptide (e.g. prolactin, placental lactogen or porcine growth  
 CC hormone) which has at least 15% homology with hGH alpha-carbon  
 CC coordinates within about 2-3.5 angstroms of hGH alpha-carbon coordinates  
 CC for about 60% of the analogue sequence, where any interaction of the  
 CC analogue with the target is different from target interaction with hGH;  
 CC (b) substituting DNA encoding an analogous polypeptide segment from the  
 CC analogue into DNA encoding the full length hGH, and expressing a  
 CC segment-substituted polypeptide; (c) contacting the segment-substituted  
 CC polypeptide with the target to determine interaction; (d) repeating steps  
 CC (b) and (c) with a second analogous polypeptide segment; and  
 CC (e) comparing the difference between activity of the first and second  
 CC segment-substituted polypeptides as an indication of the location of  
 CC the unknown active domain in hGH. The method is useful for determining  
 CC the relationship between structure and function of known polypeptide  
 CC sequences. The present sequence is that of human growth hormone  
 CC mutant substituted with residues from an hGH analogue (prolactin,  
 CC placental lactogen or porcine growth hormone).  
 CC Note: The present sequence is not shown in the specification but was  
 CC created by the indexer using the mature hGH sequence and information  
 CC contained in the specification.  
 XX  
 SQ Sequence 191 AA;

Query Match 58.1%; Score 463; DB 23; Length 191;  
 Best Local Similarity 71.0%; Pred. No. 7.1e-22;  
 Matches 103; Conservative 8; Mismatches 18; Indels 16; Gaps 5;  
 QY 2 FPTIPLSRFDNAMLRAHRLHQLAFDYOEFEEAYIPKEOKYSFLQNPOTSLSFSESPT 61  
 Db 1 FPTIPLSRFDNAMLRAHRLHQLAFDYOEFEEAYIPKEOKYSFLQNPOTSLSFSESPT 60  
 QY 62 PSNREETQOKSNLELLRISLLLIQSMLPEVQLGTGPRFVNOHLCGS-----HLV-----E 111  
 Db 61 PSNREETQOKSNLELLRISLLLIQSMLPEVQLGTGPRFVNOHLCGS-----HLV-----E 111  
 QY 112 ALYLVCG--ERGFFYTPKTRGIVEQ 134  
 Db 120 GIQTLMGRLDG----SPRTGQIFKQ 141

RESULT 14

ABG94977

ID ABG94977 standard; Protein: 191 AA.

XX AC ABG94977;

XX 03-DEC-2002 (first entry)

XX Human growth hormone mutant Y103A.

XX Growth hormone: placental lactogen; prolactin; active domain; hGH;  
 KW structure-function relationship; segment-substituted polypeptide;  
 KW mutant; mutein.  
 XX

OS Homo sapiens.

OS Synthetic.  
 XX US6428954-B1.  
 XX 06-AUG-2002.  
 XX 06-JUN-1995; 95US-0483039.  
 XX 26-OCT-1989; 89US-0428066.  
 PR 27-APR-1992; 92US-0875204.  
 PR 13-OCT-1992; 92US-0560227.  
 PR 02-FEB-1994; 94US-0190723.  
 PR 28-OCT-1998; 88US-0264611.  
 XX (GETH ) GENENTECH INC.  
 XX Wells JA. Cunningham BC.  
 WPI: 2002-696875/75.  
 PT Identifying active domains within cloned polypeptides of known amino  
 PT acid sequence by substituting analog segments into the parent  
 PT polypeptide is useful to determine the relationship between structure  
 PT and function.  
 XX  
 PS Example 1; Page -: 86pp; English.  
 XX The invention relates to identifying an unknown active domain in a region  
 CC of known amino acid sequence in a parent polypeptide e.g. human growth  
 CC hormone (hGH) which has been cloned and has a pre-identified biological  
 CC activity, where the active domain interacts with a target when the parent  
 CC polypeptide is in its native-folded form and the interaction is  
 CC responsible for the biological activity comprising: (a) comparing the  
 CC amino acid sequence or polypeptide structure in the region of known amino  
 CC acid sequence of hGH with the amino acid sequence of an analogue  
 CC polypeptide (e.g. prolactin, placental lactogen or porcine growth  
 CC hormone) which has at least 15% homology with hGH alpha-carbon  
 CC coordinates within about 2-3.5 angstroms of hGH alpha-carbon coordinates  
 CC for about 60% of the analogue sequence, where any interaction of the  
 CC analogue with the target is different from target interaction with hGH;  
 CC (b) substituting DNA encoding an analogous polypeptide segment from the  
 CC analogue into DNA encoding the full length hGH, and expressing a  
 CC segment-substituted polypeptide; (c) contacting the segment-substituted  
 CC polypeptide with the target to determine interaction; (d) repeating steps  
 CC (b) and (c) with a second analogous polypeptide segment; and  
 CC (e) comparing the difference between activity of the first and second  
 CC segment-substituted polypeptides as an indication of the location of  
 CC the unknown active domain in hGH. The method is useful for determining  
 CC the relationship between structure and function of known polypeptide  
 CC sequences. The present sequence is that of human growth hormone  
 CC mutant substituted at functionally important residues and used in  
 CC the method of the invention.  
 CC Note: The present sequence is not shown in the specification but was  
 CC created by the indexer using the mature hGH sequence and information  
 CC contained in the specification.  
 XX  
 SQ Sequence 191 AA;

Query Match 58.1%; Score 463; DB 23; Length 191;

Best Local Similarity 70.3%; Pred. No. 7.1e-22;

Matches 102; Conservative 7; Mismatches 20; Indels 16; Gaps 4;

QY 2 FPTIPLSRFDNAMLRAHRLHQLAFDYOEFEEAYIPKEOKYSFLQNPOTSLSFSESPT 61

Db 1 FPTIPLSRFDNAMLRAHRLHQLAFDYOEFEEAYIPKEOKYSFLQNPOTSLSFSESPT 60

QY 62 PSNREETQOKSNLELLRISLLLIQSMLPEVQLGTGPRFVNOHLCGS-----HLVE 111  
 Db 61 PSNREETQOKSNLELLRISLLLIQSMLPEVQLGTGPRFVNOHLCGS-----HLVE 111

QY 112 ALYLVCG--ERGFFYTPKTRGIVEQ 134



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2003, 12:37:10 : Search time 23.93 seconds  
(without alignments)  
265.217 Million cell updates/sec

Title: us-09-423-100-7

Perfect score: 797

Sequence: 1 MFPTPLSLRFDNAMLRAHR.....IVFOCCTSI(SLY)QLENYCN 150

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA\*

1: /cgn2\_6/ptodata/1/1aa/5A-COMB.pcp:\*  
2: /cgn2\_6/ptodata/1/1aa/5B-COMB.pcp:\*  
3: /cgn2\_6/ptodata/1/1aa/6A-COMB.pcp:\*  
4: /cgn2\_6/ptodata/1/1aa/6B-COMB.pcp:\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS-COMB.pcp:\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pcp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score if the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	466	58.5	192	1	Sequence 1, Appl
2	461	57.8	191	4	Sequence 5, Appl
3	461	57.8	194	2	Sequence 4, Appl
4	461	57.8	194	3	Sequence 4, Appl
5	461	57.8	217	3	Sequence 10, Appl
6	461	57.8	217	3	Sequence 10, Appl
7	461	57.8	217	3	Sequence 10, Appl
8	461	57.8	217	3	Sequence 11, Appl
9	461	57.8	217	3	Sequence 11, Appl
10	461	57.8	241	4	Sequence 25, Appl
11	461	57.8	245	4	Sequence 66, Appl
12	461	57.8	274	3	Sequence 71, Appl
13	461	57.8	360	3	Sequence 73, Appl
14	455	57.1	191	4	Sequence 1, Appl
15	455	57.1	217	1	Sequence 4, Appl
16	455	57.1	217	1	Sequence 4, Appl
17	455	57.1	217	2	Sequence 51, Appl
18	455	57.1	217	2	Sequence 51, Appl
19	455	57.1	217	4	Sequence 4, Appl
20	454	57.0	400	4	Sequence 37, Appl
21	454	57.0	401	4	Sequence 36, Appl
22	448	56.2	191	3	Sequence 18, Appl
23	446	56.0	191	3	Sequence 16, Appl
24	446	56.0	191	3	Sequence 20, Appl
25	365.5	45.9	176	3	Sequence 1, Appl
26	365.5	45.9	176	4	Sequence 1, Appl
27	359.5	45.1	176	3	Sequence 2, Appl

Sequence 2, Appl  
Patent No. 5424199  
Sequence 5, Appl  
Sequence 5, Appl  
Sequence 5, Appl  
Sequence 8, Appl  
Sequence 39, Appl  
Sequence 39, Appl  
Sequence 48, Appl  
Sequence 48, Appl  
Sequence 45, Appl  
Sequence 45, Appl  
Sequence 8, Appl  
Sequence 1, Appl  
Sequence 4, Appl  
Sequence 6, Appl  
Sequence 2, Appl

28 359.5 45.1 176 4 US-08-990-774-2  
29 343 43.0 168 6 5424199-3  
30 334.5 42.0 198 1 US-08-187-756C-5  
31 334.5 42.0 198 2 US-08-710-324A-5  
32 334.5 42.0 198 4 US-09-411-657-5  
33 334.5 39.7 96 1 US-08-160-376A-5  
34 316.5 39.7 96 1 US-08-389-487-8  
35 314 39.4 137 1 US-08-400-256-39  
36 314 39.4 137 3 US-08-975-365-39  
37 309.5 38.8 146 1 US-08-400-256-48  
38 309.5 38.8 146 3 US-08-975-365-48  
39 309 38.8 145 1 US-08-400-256-45  
40 309 38.8 145 3 US-08-975-365-45  
41 306.5 38.5 191 1 US-08-468-824-8  
42 306.5 38.5 216 2 US-09-105-651-1  
43 304.5 38.2 191 1 US-07-963-331D-4  
44 304 38.1 163 1 US-08-160-376A-6  
45 301.5 37.8 190 1 US-08-388-267C-2

#### ALIGNMENTS

#### RESULT 1

US-08-093-383-1  
Sequence 1, Application US/08093383  
Patent No. 5489529  
GENERAL INFORMATION:  
APPLICANT: DeBoer, Herman A.  
APPLICANT: Heyneker, Herbert L.  
APPLICANT: Seeburg, Peter H.  
TITLE OF INVENTION: DNA for Expression of Bovine Growth Hormone  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESS: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080

COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patlin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/093383  
FILING DATE: 14-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/619827  
FILING DATE: 28-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/198824  
FILING DATE: 05-APR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 06/632361  
FILING DATE: 19-JUL-1984  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 06/303687  
FILING DATE: 18-SEP-1981  
ATTORNEY/AGENT INFORMATION:  
NAME: Johnston, Sean A.  
REGISTRATION NUMBER: P35,910  
REFERENCE/DOCKET NUMBER: 46C4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-3562  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 192 amino acids  
TYPE: amino acid

```
; TOPOLOGY: linear
US-08-093-383-1

Query Match      58.5%  Score 466;  DB 1;  Length 192;
Best Local Similarity 70.3%;  Pred. No. 1.le-42;
Matches 103;  Conservative 7;  Mismatches 20;  Indels 16;  Gaps 4;

QY  1 MFPTPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPOTSLSFSESIP 60
DB  1 MFPTPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPOTSLSFSESIP 60

QY  61 TPSNREETQOKSNLELLRISLLLIQSWLEPVOLGTGPRFVNOHLGCS-----HLV 110
DB  61 TPSNREETQOKSNLELLRISLLLIQSWLEPVQF-LRSVFANSLSVYGASDSNVYDLKDL 119

QY  111 EALYLVCG--ERGFYTPKTRGIVEQ 134
DB  120 EGIOTLMGRLEDG---SPRTGQIFKQ 142

RESULT 2
US-09-284-878-5
; Sequence 5, Application US/09284878
; Patent No. 6342375
; GENERAL INFORMATION:
; APPLICANT: Olazaran, Martha Guerrero
; APPLICANT: Saldaña, Hugo Barrera
; APPLICANT: Salvado, Jose Maria Viader
; TITLE OF INVENTION: Genetically Modified Methylothrophic P. pastoris Yeast for the
; TITLE OF INVENTION: Production and Secretion of the Human Growth Hormone
; FILE REFERENCE: 1829.0010000
; CURRENT APPLICATION NUMBER: US/09/284.878
; CURRENT FILING DATE: 1999-07-21
; PRIOR APPLICATION NUMBER: PCT/MX97/00033
; PRIOR FILING DATE: 1997-10-24
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-284-878-5

Query Match      57.8%  Score 461;  DB 4;  Length 191;
Best Local Similarity 70.3%;  Pred. No. 3.7e-42;
Matches 102;  Conservative 7;  Mismatches 20;  Indels 16;  Gaps 4;

QY  2 FPTPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPOTSLSFSESIP 61
DB  1 FPTPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPOTSLSFSESIP 60

QY  62 PSNREETQOKSNLELLRISLLLIQSWLEPVOLGTGPRFVNOHLGCS-----HLVE 111
DB  61 PSNREETQOKSNLELLRISLLLIQSWLEPVQF-LRSVFANSLSVYGASDSNVYDLKDL 119

QY  112 ALYLVCG--ERGFYTPKTRGIVEQ 134
DB  120 GIOTLMGRLEDG---SPRTGQIFKQ 141

RESULT 3
US-08-383-621-4
; Sequence 4, Application US/08383621
; Patent No. 5951972
; GENERAL INFORMATION:
; APPLICANT: Daley, Michael J.
; APPLICANT: Buckwalter, Brian L.
; APPLICANT: Cady, Susan M.
; APPLICANT: Shieh, Hong-Ming
; APPLICANT: Bohlen, Peter
; APPLICANT: Seddon, Andrew P.
; TITLE OF INVENTION: Stabilization of Somatotropins And Other
; TITLE OF INVENTION: Proteins By Modification Of Cysteine Residues
```

```
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dr. Estelle J. Tsevdos
; STREET: 1937 West Main Street, P.O. Box 60
; CITY: Stamford
; STATE: Connecticut
; COUNTRY: U.S.A.
; ZIP: 06904-0060
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/383,621
; FILING DATE: 06-FEB-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/766,142
; FILING DATE: 25-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsevdos, Estelle J.
; REGISTRATION NUMBER: 31,145
; REFERENCE/DOCKET NUMBER: 31,278-91
; TELEPHONE: 203-321-2756
; TELEFAX: 203-321-2971
; TELEX: 203-710-474-4059
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 194 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-383-621-4

Query Match      57.8%  Score 461;  DB 2;  Length 194;
Best Local Similarity 70.3%;  Pred. No. 3.8e-42;
Matches 102;  Conservative 7;  Mismatches 20;  Indels 16;  Gaps 4;

QY  2 FPTPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPOTSLSFSESIP 61
DB  4 FPTPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPOTSLSFSESIP 63

QY  62 PSNREETQOKSNLELLRISLLLIQSWLEPVOLGTGPRFVNOHLGCS-----HLVE 111
DB  54 PSNREETQOKSNLELLRISLLLIQSWLEPVQF-LRSVFANSLSVYGASDSNVYDLKDL 122

QY  112 ALYLVCG--ERGFYTPKTRGIVEQ 134
DB  123 GIOTLMGRLEDG---SPRTGQIFKQ 144

RESULT 4
US-08-459-906-4
; Sequence 4, Application US/08459906
; Patent No. 6010995
; GENERAL INFORMATION:
; APPLICANT: Daley, Michael J.
; APPLICANT: Buckwalter, Brian L.
; APPLICANT: Cady, Susan M.
; APPLICANT: Shieh, Hong-Ming
; APPLICANT: Bohlen, Peter
; APPLICANT: Seddon, Andrew P.
; TITLE OF INVENTION: Stabilization of Somatotropins and Other
; TITLE OF INVENTION: Proteins By Modification of Cysteine Residues
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: U.S.A.
```







OTHER INFORMATION: Description of Artificial Sequence: Designated as  
OTHER INFORMATION: an amino acid sequence of MWPsp-MWPmp20-TEV-C-GH  
US-09-280-030-66

Query Match 57.8%; Score 461; DB 4; Length 245;  
Best Local Similarity 70.3%; Pred. No. 5.2e-42;  
Matches 102; Conservative 7; Mismatches 20; Indels 16; Gaps 4;  
QY 2 FPTPLSLFONAMLRHRLHQLAFDTYQEEFAYIPKEQKYSFLQNPQTSLSFSESPT 61  
DB 55 FPTPLSLFONAMLRHRLHQLAFDTYQEEFAYIPKEQKYSFLQNPQTSLSFSESPT 114  
QY 62 PSNREETOQKSNLELRISLLIQSWLEPVQGTGRFVNOHLGCS-----HLVE 111  
DB 115 PSNREETOQKSNLELRISLLIQSWLEPVQGTGRFVNOHLGCS-----HLVE 173  
QY 112 ALYLVCG--ERGFFYTPKTRGIVEQ 134  
DB 174 GIOTLMGRLEDG---SPRTGQIFKQ 195

RESULT 12  
US-08-784-582-71  
Sequence 71, Application US/08784582  
Patent No. 6110707  
GENERAL INFORMATION:  
APPLICANT: Newgard, Christopher B.  
APPLICANT: Halban, Philippe A.  
APPLICANT: No. 6110707minq, Karl D.  
APPLICANT: Clark, Samuel A.  
APPLICANT: Thigpen, Anice E.  
APPLICANT: Quade, Christian  
APPLICANT: Kruse, Fred  
APPLICANT: McGarry, Dennis  
TITLE OF INVENTION: RECOMBINANT EXPRESSION OF PROTEINS FROM  
NUMBER OF SEQUENCES: 79  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210  
COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/784,582  
FILING DATE: Concurrently Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION NUMBER: US 60/028,427  
FILING DATE: 15-OCT-1996  
APPLICATION NUMBER: US 08/589,028  
FILING DATE: 19-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Highlander, Steven L.  
REGISTRATION NUMBER: 37,642  
REFERENCE/DOCKET NUMBER: UTSD:514  
TELEPHONE: 512/418-3000  
TELEFAX: 512/474-7577  
INFORMATION FOR SEQ ID NO: 71:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 274 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear

US-08-784-582-71  
Query Match 57.8%; Score 461; DB 3; Length 360;  
Best Local Similarity 70.3%; Pred. No. 8.5e-42;  
Matches 102; Conservative 7; Mismatches 20; Indels 16; Gaps 4;  
QY 2 FPTPLSLFONAMLRHRLHQLAFDTYQEEFAYIPKEQKYSFLQNPQTSLSFSESPT 61  
DB 55 FPTPLSLFONAMLRHRLHQLAFDTYQEEFAYIPKEQKYSFLQNPQTSLSFSESPT 114  
QY 62 PSNREETOQKSNLELRISLLIQSWLEPVQGTGRFVNOHLGCS-----HLVE 111  
DB 115 PSNREETOQKSNLELRISLLIQSWLEPVQGTGRFVNOHLGCS-----HLVE 173  
QY 112 ALYLVCG--ERGFFYTPKTRGIVEQ 134  
DB 174 GIOTLMGRLEDG---SPRTGQIFKQ 195

Query Match 57.8%; Score 461; DB 3; Length 274;  
Best Local Similarity 70.3%; Pred. No. 6e-42;  
Matches 102; Conservative 7; Mismatches 20; Indels 16; Gaps 4;

QY 2 FPTPLSLFONAMLRHRLHQLAFDTYQEEFAYIPKEQKYSFLQNPQTSLSFSESPT 61  
DB 55 FPTPLSLFONAMLRHRLHQLAFDTYQEEFAYIPKEQKYSFLQNPQTSLSFSESPT 114  
QY 62 PSNREETOQKSNLELRISLLIQSWLEPVQGTGRFVNOHLGCS-----HLVE 111  
DB 87 PSNREETOQKSNLELRISLLIQSWLEPVQGTGRFVNOHLGCS-----HLVE 145  
QY 112 ALYLVCG--ERGFFYTPKTRGIVEQ 134  
DB 146 GIOTLMGRLEDG---SPRTGQIFKQ 167

RESULT 13  
US-08-784-582-73  
Sequence 73, Application US/08784582  
Patent No. 6110707  
GENERAL INFORMATION:  
APPLICANT: Newgard, Christopher B.  
APPLICANT: Halban, Philippe A.  
APPLICANT: No. 6110707minq, Karl D.  
APPLICANT: Clark, Samuel A.  
APPLICANT: Thigpen, Anice E.  
APPLICANT: Quade, Christian  
APPLICANT: Kruse, Fred  
APPLICANT: McGarry, Dennis  
TITLE OF INVENTION: RECOMBINANT EXPRESSION OF PROTEINS FROM  
NUMBER OF SEQUENCES: 79  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210  
COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/784,582  
FILING DATE: Concurrently Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION NUMBER: US 60/028,427  
FILING DATE: 15-OCT-1996  
APPLICATION NUMBER: US 08/589,028  
FILING DATE: 19-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Highlander, Steven L.  
REGISTRATION NUMBER: 37,642  
REFERENCE/DOCKET NUMBER: UTSD:514  
TELEPHONE: 512/418-3000  
TELEFAX: 512/474-7577  
INFORMATION FOR SEQ ID NO: 73:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 360 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear

US-08-784-582-73  
Query Match 57.8%; Score 461; DB 3; Length 360;  
Best Local Similarity 70.3%; Pred. No. 8.5e-42;  
Matches 102; Conservative 7; Mismatches 20; Indels 16; Gaps 4;  
QY 2 FPTPLSLFONAMLRHRLHQLAFDTYQEEFAYIPKEQKYSFLQNPQTSLSFSESPT 61  
DB 55 FPTPLSLFONAMLRHRLHQLAFDTYQEEFAYIPKEQKYSFLQNPQTSLSFSESPT 114  
QY 62 PSNREETOQKSNLELRISLLIQSWLEPVQGTGRFVNOHLGCS-----HLVE 111  
DB 87 PSNREETOQKSNLELRISLLIQSWLEPVQGTGRFVNOHLGCS-----HLVE 145  
QY 112 ALYLVCG--ERGFFYTPKTRGIVEQ 134  
DB 146 GIOTLMGRLEDG---SPRTGQIFKQ 167

US-08-784-582-73  
Query Match 57.8%; Score 461; DB 3; Length 360;  
Best Local Similarity 70.3%; Pred. No. 8.5e-42;  
Matches 102; Conservative 7; Mismatches 20; Indels 16; Gaps 4;

Matches 102; Conservative 7; Mismatch 20; Indels 16; Gaps 4;  
QY 2 FPTPLSRFDNAMLRAHRLHQLAFDTYQEEFAYIPKEOKYSFLONPQTSLSFSSEIPT 61  
Db 27 FPTPLSRFDNAMLRAHRLHQLAFDTYQEEFAYIPKEOKYSFLONPQTSLSFSSEIPT 86  
QY 62 PSNREETQOKSNLELLRISLLLSQSWLEPVOLGTGRFVNOHLGGS-----HLVE 111  
Db 87 PSNREETQOKSNLELLRISLLLSQSWLEPVQF-LRSVFANSLSVYGASDSNVYDLKDLFE 145  
QY 112 ALYLVCG--ERGFFYTPKTRGIVEQ 134  
Db 146 GIOTLMGRLEDG---SPRTGQIFKQ 167

## RESULT 14

US-09-465-461-1  
; Sequence 1, Application US/09465461  
; Patent No. 6348444  
; GENERAL INFORMATION:  
; APPLICANT: C APPEL, Scott  
; TITLE OF INVENTION: Human Growth Hormone to stimulate hematopoiesis and immune reconst  
; TITLE OF INVENTION: after hematopoietic stem cell transplantation in humans  
; FILE REFERENCE: CHAPPEL-6.1  
; CURRENT APPLICATION NUMBER: US/09/465.461  
; PRIOR FILING DATE: 1999-12-17  
; PRIOR APPLICATION NUMBER: 60/112,668  
; PRIOR FILING DATE: 1998-12-17  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 191  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-09-465-461-1

Query Match 57.1%; Score 455; DB 4; Length 191;  
Best Local Similarity 69.7%; Pred. No. 1.7e-41;

Matches 101; Conservative 7; Mismatch 21; Indels 16; Gaps 4;  
QY 2 FPTPLSRFDNAMLRAHRLHQLAFDTYQEEFAYIPKEOKYSFLONPQTSLSFSSEIPT 61  
Db 1 FPTPLSRFDNAMLRAHRLHQLAFDTYQEEFAYIPKEOKYSFLONPQTSLSFSSEIPT 60  
QY 62 PSNREETQOKSNLELLRISLLLSQSWLEPVOLGTGRFVNOHLGGS-----HLVE 111  
Db 61 PSNREETQOKSNLELLRISLLLSQSWLEPVQF-LRSVFANSLSVYGASDSNVYDLKDLFE 179  
QY 112 ALYLVCG--ERGFFYTPKTRGIVEQ 134  
Db 120 GIOTLMGRLEDG---SPRTGQIFKQ 141

## RESULT 15

US-08-187-756C-4  
; Sequence 4, Application US/08187756C  
; Patent No. 5597709  
; GENERAL INFORMATION:  
; APPLICANT: ROSEN, ET AL.  
; TITLE OF INVENTION: Human Growth Hormone  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAYN, GILFILLAN,  
; ADDRESS: CECCHI, STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS

SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/187.756C  
FILING DATE: January 27, 1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: FERRARO, GREGORY D.  
REGISTRATION NUMBER: 36,114  
REFERENCE/DOCKET NUMBER: 325800-55  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 217 AMINO ACIDS  
TYPE: AMINO ACID  
STRANDEDNESS:  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
US-08-187-756C-4  
Query Match 57.1%; Score 455; DB 1; Length 217;  
Best Local Similarity 69.7%; Pred. No. 2e-41;  
Matches 101; Conservative 7; Mismatch 21; Indels 16; Gaps 4;  
QY 2 FPTPLSRFDNAMLRAHRLHQLAFDTYQEEFAYIPKEOKYSFLONPQTSLSFSSEIPT 61  
Db 27 FPTPLSRFDNAMLRAHRLHQLAFDTYQEEFAYIPKEOKYSFLONPQTSLSFSSEIPT 86  
QY 62 PSNREETQOKSNLELLRISLLLSQSWLEPVOLGTGRFVNOHLGGS-----HLVE 111  
Db 87 PSNREETQOKSNLELLRISLLLSQSWLEPVQF-LRSVFANSLSVYGASDSNVYDLKDLFE 145  
QY 112 ALYLVCG--ERGFFYTPKTRGIVEQ 134  
Db 146 GIOTLMGRLEDG---SPRTGQIFKQ 167

Search completed: September 16, 2003, 12:41:25  
Job time : 24.93 sec's

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2003, 12:38:55 : Search time 47.2763 seconds  
(without alignments)  
472.415 Million cell updates/sec

**Title:** US-09-423-100-7

**Perfect score:** 797

**Sequence:** 1 MFPTPLSLRLEFONAMLAHR.....IVQCCCTSLCSYQLENVCN 150

**Scoring table:** BLOSUM62

Gapop 10.0 , Gapext 0.5

**Searched:** 556269 seqs, 14893369 residues

Total number of hits satisfying chosen parameters: 556269

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

**Database :** Published Applications AA:\*

```
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	797	100.0	150	14	US-10-054-873-7
2	555.5	69.7	107	14	US-10-054-873-6
3	470	59.0	92	14	US-10-054-873-2
4	461	57.8	191	11	US-09-984-010-23
5	461	57.8	191	12	US-10-153-207-1
6	461	57.8	191	12	US-10-400-377-1
7	461	57.8	191	12	US-10-400-708-1
8	461	57.8	191	12	US-10-298-148-1
9	461	57.8	214	12	US-10-153-207-6
10	461	57.8	217	9	US-09-929-918-9
11	461	57.8	245	9	US-09-380-030-66
12	455	57.1	217	9	US-09-853-688-2
13	455	57.1	217	12	US-09-969-748C-4
14	454	57.0	217	10	US-09-804-409A-16
15	448	56.2	217	9	US-09-853-688-4

16	445	55.8	191	12	US-09-824-200-12	Sequence 12, Appl
17	440	55.2	191	12	US-10-300-822-1	Sequence 1, Appl
18	433	54.3	191	12	US-10-289-845-3	Sequence 3, Appl
19	399	50.1	217	9	US-09-850-887-3	Sequence 3, Appl
20	381	47.8	163	15	US-10-043-487-350	Sequence 350, App
21	381	47.8	191	12	US-10-153-207-2	Sequence 2, Appl
22	373	46.8	229	15	US-10-103-313-411	Sequence 411, App
23	345	43.3	246	15	US-10-188-246-18	Sequence 18, Appl
24	301.5	37.8	190	12	US-10-153-207-3	Sequence 3, Appl
25	294.5	37.0	138	10	US-09-861-687-19	Sequence 19, Appl
26	294	36.9	52	14	US-10-054-873-5	Sequence 5, Appl
27	286	35.9	147	9	US-09-736-611-8	Sequence 8, Appl
28	286	35.9	147	9	US-09-740-359-7	Sequence 7, Appl
29	284.5	35.7	124	9	US-09-894-711-18	Sequence 18, Appl
30	284.5	35.7	144	9	US-09-736-611-6	Sequence 6, Appl
31	284.5	35.7	144	9	US-09-740-359-5	Sequence 5, Appl
32	284.5	35.7	146	10	US-09-894-711-5	Sequence 5, Appl
33	278.5	34.9	51	11	US-09-858-935B-5	Sequence 5, Appl
34	278.5	34.9	51	14	US-10-028-410-3	Sequence 3, Appl
35	278	34.9	117	9	US-09-280-030-63	Sequence 63, Appl
36	277	34.8	96	10	US-09-947-563-4	Sequence 4, Appl
37	275.5	34.6	124	9	US-09-736-611-12	Sequence 12, Appl
38	275.5	34.6	124	9	US-09-740-359-12	Sequence 12, Appl
39	275.5	34.6	124	10	US-09-894-711-12	Sequence 12, Appl
40	275.5	34.6	125	9	US-09-736-611-10	Sequence 10, Appl
41	275.5	34.6	125	9	US-09-740-359-10	Sequence 10, Appl
42	275.5	34.6	125	10	US-09-894-711-10	Sequence 10, Appl
43	273	34.3	50	14	US-10-066-009A-3	Sequence 3, Appl
44	271	34.0	96	10	US-09-947-563-5	Sequence 5, Appl
45	270	33.9	110	9	US-09-205-658-125	Sequence 125, App

#### ALIGNMENTS

#### RESULT 1

```
US-10-054-873-7
: Sequence 7, Application US/10054873
: Publication No. US20020164712A1
: GENERAL INFORMATION:
: APPLICANT: Gan, Zhong Ru
: TITLE OF INVENTION: Chimeric Protein Containing an
: Intramolecular Chapterone-Like Sequence
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSER: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor
: CITY: San Francisco,
: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/10/054,873
: FILING DATE: 22-Jan-2002
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO PCT/CN98/00052
: FILING DATE: 31-MAR-1998
: APPLICATION NUMBER: US 09/423,100
: FILING DATE: 11-DEC-2000
: ATTORNEY/AGENT INFORMATION:
: NAME: Mycroft, Frank J
: REGISTRATION NUMBER: 46,946
: REFERENCE/DOCKET NUMBER: 020167-00013005
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 150 amino acids
: TYPE: amino acid
```

```

; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-10-054-873-7

Query Match          100.0%; Score 797; DB 14; Length 150;
Best Local Similarity 100.0%; Pred. No. 1.4e-80;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFPTPLSLFDNAMLRAHRLHQLAFDFTYQEEFAYIPKEQKYSFLQNPQTSLSFSSESIP 60
DB 1 MFPTPLSLFDNAMLRAHRLHQLAFDFTYQEEFAYIPKEQKYSFLQNPQTSLSFSSESIP 60
QY 61 TPSNREETQOKSNLELLRISLLLIQSWLEPVOLGTGPRFVNOHLCGSHLVEALYLVCGER 120
DB 61 TPSNREETQOKSNLELLRISLLLIQSWLEPVOLGTGPRFVNOHLCGSHLVEALYLVCGER 120
QY 121 GFFYTPKTRGIVEQCCTSIQSLYQLENYCN 150
DB 121 GFFYTPKTRGIVEQCCTSIQSLYQLENYCN 150

RESULT 2
US-10-054-873-6
; Sequence 6, Application US/10054873
; Publication No. US20020164712A1
; GENERAL INFORMATION:
; APPLICANT: Gan, Zhong Ru
; TITLE OF INVENTION: Chimeric Protein Containing an
; Intramolecular Chaperone-Like Sequence
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/054,873
; FILING DATE: 22-Jan-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/CN98/00052
; FILING DATE: 31-MAR-1998
; APPLICATION NUMBER: 39/423,100
; FILING DATE: 11-DEC-2000
; ATTORNEY/AGENT INFORMATION:
; NAME: Mycroft, Frank J.
; REGISTRATION NUMBER: 46,946
; REFERENCE/DOCKET NUMBER: 020167-000130US
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 107 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-054-873-6

Query Match          69.7%; Score 555.5; DB 14; Length 107;
Best Local Similarity 71.3%; Pred. No. 5.9e-54;
Matches 107; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MFPTPLSLFDNAMLRAHRLHQLAFDFTYQEEFAYIPKEQKYSFLQNPQTSLSFSSESIP 60
DB 1 MFPTPLSLFDNAMLRAHRLHQLAFDFTYQEEFAYIPKEQKYSFLQNPQTSLSFSSESIP 60
QY 61 TPSNREETQOKSNLELLRISLLLIQSWLEPVOLGTGPRFVNOHLCGSHLVEALYLVCGER 120
DB 61 TPSNREETQOKSNLELLRISLLLIQSWLEPVOLGTGPRFVNOHLCGSHLVEALYLVCGER 120
QY 121 GFFYTPKTRGIVEQCCTSIQSLYQLENYCN 150
DB 121 GFFYTPKTRGIVEQCCTSIQSLYQLENYCN 150

RESULT 3
US-10-054-873-2
; Sequence 2, Application US/10054873
; Publication No. US20020164712A1
; GENERAL INFORMATION:
; APPLICANT: Gan, Zhong Ru
; TITLE OF INVENTION: Chimeric Protein Containing an
; Intramolecular Chaperone-Like Sequence
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/054,873
; FILING DATE: 22-Jan-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/CN98/00052
; FILING DATE: 31-MAR-1998
; APPLICATION NUMBER: 39/423,100
; FILING DATE: 11-DEC-2000
; ATTORNEY/AGENT INFORMATION:
; NAME: Mycroft, Frank J.
; REGISTRATION NUMBER: 46,946
; REFERENCE/DOCKET NUMBER: 020167-000130US
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 92 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-054-873-2

Query Match          59.0%; Score 470; DB 14; Length 92;
Best Local Similarity 100.0%; Pred. No. 1.5e-44;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFPTPLSLFDNAMLRAHRLHQLAFDFTYQEEFAYIPKEQKYSFLQNPQTSLSFSSESIP 60
DB 1 MFPTPLSLFDNAMLRAHRLHQLAFDFTYQEEFAYIPKEQKYSFLQNPQTSLSFSSESIP 60
QY 61 TPSNREETQOKSNLELLRISLLLIQSWLEPVOLGTGPRFVNOHLCGSHLVEALYLVCGER 120
DB 61 TPSNREETQOKSNLELLRISLLLIQSWLEPVOLGTGPRFVNOHLCGSHLVEALYLVCGER 120
QY 121 GFFYTPKTRGIVEQCCTSIQSLYQLENYCN 150
DB 121 GFFYTPKTRGIVEQCCTSIQSLYQLENYCN 150

RESULT 4
US-09-984-010-23
; Sequence 23, Application US/09984010
; Publication No. US20030104578A1
; GENERAL INFORMATION:

```

APPLICANT: Ballance, David James  
TITLE OF INVENTION: RECOMBINANT FUSION PROTEINS TO GROWTH HORMONE  
AND SERUM ALBUMIN  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FINNegan, Henderson, Farabow, Garrett & Dunner, L.L.P.  
STREET: 1300 I Street, NW  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/984,010  
FILING DATE: 21-May-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/091,873  
FILING DATE: 25-JUN-1998  
APPLICATION NUMBER: PCI/GB96/03164  
FILING DATE: 19-DEC-1996  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 191 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
SEQUENCE DESCRIPTION: SEQ ID NO: 23:  
US-09-984-010-23

Query Match 57.8%; Score 461; DB 11; Length 191;  
Best Local Similarity 70.3%; Pred. No. 3.9e-43;  
Matches 102; Conservative 7; Mismatches 20; Indels 16; Gaps 4;  
QY 2 FPTPLSRFLFDNAMLRAHRLHQLAFDTYQEEAYIPKEQKYSFLQNPTSLSFSESIPT 61  
Db 1 FPTPLSRFLFDNAMLRAHRLHQLAFDTYQEEAYIPKEQKYSFLQNPTSLSFSESIPT 60  
QY 62 PSNREETQOKSNLELLRLISLLIQSWLEPVQGTGPRFVNQHLGCS-----HLVE 111  
Db 61 PSNREETQOKSNLELLRLISLLIQSWLEPVQGTGPRFVNQHLGCS-----HLVE 111  
QY 112 ALYLVCG--ERGFYTPKTRGIVEQ 134  
Db 120 GIQTLGRLDG---SPRTGQIFKQ 141

RESULT 5  
US-10-153-207-1  
Sequence 1, Application US/10153207  
Publication No. US20030153003A1  
GENERAL INFORMATION:  
APPLICANT: James A. Wells  
TITLE OF INVENTION: GROWTH HORMONE VARIANTS  
FILE REFERENCE: 669.12-US-C7  
CURRENT APPLICATION NUMBER: US/10/153,207  
PRIOR FILING DATE: 2002-05-22  
PRIOR APPLICATION NUMBER: 08/479,884  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 08/190,723  
PRIOR FILING DATE: 1994-02-02  
PRIOR APPLICATION NUMBER: 07/960,227  
PRIOR FILING DATE: 1992-10-13  
PRIOR APPLICATION NUMBER: 07/875,204  
PRIOR FILING DATE: 1992-04-27  
PRIOR APPLICATION NUMBER: 07/428,066

PRIOR FILING DATE: 1989-10-26  
PRIOR APPLICATION NUMBER: 07/264,611  
PRIOR FILING DATE: 1988-10-28  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: FastSeq for Windows version 4.0  
SEQ ID NO 1  
LENGTH: 191  
TYPE: PRT  
ORGANISM: Homo Sapiens  
US-10-153-207-1  
Query Match 57.8%; Score 461; DB 12; Length 191;  
Best Local Similarity 70.3%; Pred. No. 3.9e-43;  
Matches 102; Conservative 7; Mismatches 20; Indels 16; Gaps 4;  
QY 2 FPTPLSRFLFDNAMLRAHRLHQLAFDTYQEEAYIPKEQKYSFLQNPTSLSFSESIPT 61  
Db 1 FPTPLSRFLFDNAMLRAHRLHQLAFDTYQEEAYIPKEQKYSFLQNPTSLSFSESIPT 60  
QY 62 PSNREETQOKSNLELLRLISLLIQSWLEPVQGTGPRFVNQHLGCS-----HLVE 111  
Db 61 PSNREETQOKSNLELLRLISLLIQSWLEPVQGTGPRFVNQHLGCS-----HLVE 111  
QY 112 ALYLVCG--ERGFYTPKTRGIVEQ 134  
Db 120 GIQTLGRLDG---SPRTGQIFKQ 141

RESULT 6  
US-10-400-377-1  
Sequence 1, Application US/1040377  
Publication No. US20030162949A1  
GENERAL INFORMATION:  
APPLICANT: Cox III, George N.  
TITLE OF INVENTION: Derivatives of Growth Hormone and Related Proteins  
FILE REFERENCE: 4152-1-PUS  
CURRENT APPLICATION NUMBER: US/10/400,377  
PRIOR FILING DATE: 2003-03-26  
PRIOR APPLICATION NUMBER: US/09/462,941  
PRIOR FILING DATE: 2000-01-14  
PRIOR APPLICATION NUMBER: 60/052,516  
PRIOR FILING DATE: 1997-07-14  
NUMBER OF SEQ ID NOS: 41  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 191  
TYPE: PRT  
ORGANISM: Homo Sapiens  
US-10-400-377-1

Query Match 57.8%; Score 461; DB 12; Length 191;  
Best Local Similarity 70.3%; Pred. No. 3.9e-43;  
Matches 102; Conservative 7; Mismatches 20; Indels 16; Gaps 4;  
QY 2 FPTPLSRFLFDNAMLRAHRLHQLAFDTYQEEAYIPKEQKYSFLQNPTSLSFSESIPT 61  
Db 1 FPTPLSRFLFDNAMLRAHRLHQLAFDTYQEEAYIPKEQKYSFLQNPTSLSFSESIPT 60  
QY 62 PSNREETQOKSNLELLRLISLLIQSWLEPVQGTGPRFVNQHLGCS-----HLVE 111  
Db 61 PSNREETQOKSNLELLRLISLLIQSWLEPVQGTGPRFVNQHLGCS-----HLVE 111  
QY 112 ALYLVCG--ERGFYTPKTRGIVEQ 134  
Db 120 GIQTLGRLDG---SPRTGQIFKQ 141

RESULT 7  
US-10-400-708-1  
Sequence 1, Application US/10400708  
Publication No. US20030166865A1  
GENERAL INFORMATION:

APPLICANT: Cox III, George N  
APPLICANT: Bolder Biotechnology, Inc.  
TITLE OF INVENTION: Derivatives of Growth Hormone and Related Proteins  
FILE REFERENCE: 4152-1-PUS  
CURRENT APPLICATION NUMBER: US/10/400,708  
CURRENT FILING DATE: 2003-03-26  
PRIOR APPLICATION NUMBER: US/09/462,941  
PRIOR FILING DATE: 2000-01-14  
PRIOR APPLICATION NUMBER: 60/052,516  
PRIOR FILING DATE: 1997-07-14  
NUMBER OF SEQ ID NOS: 41  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 1  
LENGTH: 191  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-400-708-1

Query Match 57.8% Score 461; DB 12; Length 191;  
Best Local Similarity 70.3%; Pred. No. 3.9e-43;  
Matches 102; Conservative 7; Mismatches 20; Indels 16; Gaps 4;

QY 2 FPTPLSRFDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONPOTSLSESPT 61  
DB 1 FPTPLSRFDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONPOTSLSESPT 60  
QY 62 PSNREETQOKSNLELLRISLLIQSWLEPVLQGTGPRFVNHLCGS-----HLVE 111  
DB 61 PSNREETQOKSNLELLRISLLIQSWLEPVLQGTGPRFVNHLCGS-----HLVE 110  
QY 112 ALYLVCG--ERGFYTPKTRGIVEQ 134  
DB 120 GIQTLMGRLDG---SPRTQIQFKQ 141

RESULT 8  
US-10-298-148-1  
Sequence 1, Application US/10298148  
Publication No. US20030171284A1  
GENERAL INFORMATION:  
APPLICANT: Cox III, George N  
APPLICANT: Bolder Biotechnology, Inc.  
TITLE OF INVENTION: Derivatives of Growth Hormone and Related Proteins  
FILE REFERENCE: 4152-1-PUS  
CURRENT APPLICATION NUMBER: US/10/298,148  
CURRENT FILING DATE: 2002-11-15  
PRIOR APPLICATION NUMBER: US/09/462,941  
PRIOR FILING DATE: 2000-01-14  
PRIOR APPLICATION NUMBER: 60/052,516  
PRIOR FILING DATE: 1997-07-14  
NUMBER OF SEQ ID NOS: 41  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 1  
LENGTH: 191  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-298-148-1

Query Match 57.8% Score 461; DB 12; Length 191;  
Best Local Similarity 70.3%; Pred. No. 3.9e-43;  
Matches 102; Conservative 7; Mismatches 20; Indels 16; Gaps 4;

QY 2 FPTPLSRFDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONPOTSLSESPT 61  
DB 1 FPTPLSRFDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONPOTSLSESPT 60  
QY 62 PSNREETQOKSNLELLRISLLIQSWLEPVLQGTGPRFVNHLCGS-----HLVE 111  
DB 61 PSNREETQOKSNLELLRISLLIQSWLEPVLQGTGPRFVNHLCGS-----HLVE 110  
QY 112 ALYLVCG--ERGFYTPKTRGIVEQ 134  
DB 120 GIQTLMGRLDG---SPRTQIQFKQ 141

RESULT 9  
US-10-153-207-6  
Sequence 6, Application US/10153207  
Publication No. US20030153003A1  
GENERAL INFORMATION:  
APPLICANT: James A. Wells  
APPLICANT: Brian C. Cunningham  
TITLE OF INVENTION: GROWTH HORMONE VARIANTS  
FILE REFERENCE: 659.12-US-C7  
CURRENT APPLICATION NUMBER: US/10/153,207  
CURRENT FILING DATE: 2002-05-22  
PRIOR APPLICATION NUMBER: 08/479,884  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 08/190,723  
PRIOR FILING DATE: 1994-02-02  
PRIOR APPLICATION NUMBER: 07/900,227  
PRIOR FILING DATE: 1992-10-13  
PRIOR APPLICATION NUMBER: 07/875,264  
PRIOR FILING DATE: 1992-04-27  
PRIOR APPLICATION NUMBER: 07/428,066  
PRIOR FILING DATE: 1989-10-26  
PRIOR APPLICATION NUMBER: 07/264,611  
PRIOR FILING DATE: 1986-10-28  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 6  
LENGTH: 214  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-153-207-6

Query Match 57.8% Score 461; DB 12; Length 214;  
Best Local Similarity 70.3%; Pred. No. 4.5e-44;  
Matches 102; Conservative 7; Mismatches 20; Indels 15; Gaps 4;

QY 2 FPTPLSRFDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONPOTSLSESPT 61  
DB 24 FPTPLSRFDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONPOTSLSESPT 83  
QY 62 PSNREETQOKSNLELLRISLLIQSWLEPVLQGTGPRFVNHLCGS-----HLVE 111  
DB 84 PSNREETQOKSNLELLRISLLIQSWLEPVLQGTGPRFVNHLCGS-----HLVE 142  
QY 112 ALYLVCG--ERGFYTPKTRGIVEQ 134  
DB 143 GIQTLMGRLDG---SPRTQIQFKQ 164

RESULT 10  
US-09-929-918-9  
Sequence 9, Application US/09929918  
Patent No. US20020090678A1  
GENERAL INFORMATION:  
APPLICANT: Kordyum, Vitaliy A.  
APPLICANT: Chernykh, Svittlana I.  
APPLICANT: Slavchenko, Iryna Yu.  
APPLICANT: Vozianov, Oleksandr  
TITLE OF INVENTION: PHAGE-DEPENDENT SUPER PRODUCTION OF BIOLOGICALLY ACTIVE PROTEIN AND PEPTIDES  
FILE REFERENCE: PHAGE.036A  
CURRENT APPLICATION NUMBER: US/09/929,918  
CURRENT FILING DATE: 2001-08-15  
PRIOR APPLICATION NUMBER: 09/318,288  
PRIOR FILING DATE: 1999-05-25  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 9  
LENGTH: 217  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-929-918-9

Query Match 57.8% Score 461; DB 12; Length 191;  
Best Local Similarity 70.3%; Pred. No. 3.9e-43;  
Matches 102; Conservative 7; Mismatches 20; Indels 16; Gaps 4;

QY 2 FPTPLSRFDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONPOTSLSESPT 61  
DB 1 FPTPLSRFDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONPOTSLSESPT 60  
QY 62 PSNREETQOKSNLELLRISLLIQSWLEPVLQGTGPRFVNHLCGS-----HLVE 111  
DB 61 PSNREETQOKSNLELLRISLLIQSWLEPVLQGTGPRFVNHLCGS-----HLVE 110  
QY 112 ALYLVCG--ERGFYTPKTRGIVEQ 134  
DB 120 GIQTLMGRLDG---SPRTQIQFKQ 141

```

: TITLE OF INVENTION: HUMANS, THE VARIATIONS AND THEIR USES
:
: FILE REFERENCE: WCM78
:
: CURRENT APPLICATION NUMBER: US/09/853,688
:
: CURRENT FILING DATE: 2001-05-14
:
: NUMBER OF SEQ ID NOS: 56
:
: SOFTWARE: PatentIn Ver. 2.1
:
: SEQ ID NO 2
:
: LENGTH: 217
:
: TYPE: PRT
:
: ORGANISM: Homo sapiens
:
: US-09-853,688-2

```

```

Query Match      57.1%  Score 455.  DB 9.  Length. 217;
Best Local Similarity 69.7% 7; Mismatches 21; Indels 16; Gaps 4;
Matches 101; Conservative
2  FPTPLSLRFDNAMLRAHRLHQAFDITQYFFEEAYIPKEUKYSLFQNPQISLPSSESIPT 61
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
27  FPTPLSLRFDNASLRHRLHQAFDITQYFFEEAYIPKEUKYSLFQNPQISLPSSESIPT 86
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
72  PSNREETQOKSNLELLRISLLLSQSWLEPVLQGTGRFVNHQIGCS-----HIVE 111
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
87  PSNREETQOKSNLELLRISLLLSQSWLEPVLQGTGRFVNHQIGCS-----HIVE 145
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
112  ALYLVCQ--ENGFFYTPKTRILVEQ 134
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
146  GIOTLMGMLEDG-----SPRIGQIPKO 167
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||

```

```

RESULT 13
US-09-969-748C-4
: Sequence 4, Application US/09969748C
: Publication NO. US20030761809A1
: GENERAL INFORMATION:
: APPLICANT: ARIZTKE PHARMACEUTICALS, INC.
: APPLICANT: HOUSTON, Lou, L.
: APPLICANT: SHAWHAN, Philip, J.
: APPLICANT: HAWLEY, Stephen
: APPLICANT: GLYNN, Jacqueline, M.
: APPLICANT: CHAPIN, Steven
: APPLICANT: BASU, Amarash
: TITLE OF INVENTION: COMPOSITIONS AND METHODS
: FILE REFERENCE: 057920-0303
: CURRENT APPLICATION NUMBER: US/09/969-748C
: CURRENT FILING DATE: 2002-12-19
: PRIOR APPLICATION NUMBER: US 60/267,601
: PRIOR FILING DATE: 2001-02-09
: PRIOR APPLICATION NUMBER: US 60/248,914
: PRIOR FILING DATE: 2000-11-14
: PRIOR APPLICATION NUMBER: US 60/248,478
: PRIOR FILING DATE: 2000-11-13
: PRIOR APPLICATION NUMBER: US 60/237,929
: PRIOR FILING DATE: 2000-10-02
: NUMBER OF SEQ ID NOS: 115
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 4
: LENGTH: 217
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-969-748C-4

```

Query Match	57.1%	Score 455;	DB 12;	Length 217;
Best Local Similarity	69.7%	Pred. NO. 2.le-42;		
Matches 101;	Conservative 7;	Mismatches 21;	Indels 16;	Gaps 4;

  

QY	2	PPTPLSRLEDNAMLRAHRLHQLAFDTYQEFEEVAPKEQKYSFLONPQTSLSFSSEIPT	61
Db	27	PPTPLSRLEDNALSRAHRLHQLAFDTYQEFEEVAPKEQKYSFLONPQTSLSFSSEIPT	86
QY	62	PSNREETQOKSNLELLRISLLLTQSWLEPVLCTGPRFVNHLGCS-----HLVE	111
Db	87	PSNREETQOKSNLELLRISLLLTQSWLEPVLCTGPRFVNHLGCS-----HLVE	145

QY 112 ALYLVCG--ERGFFYTPKTRGIVEQ 134  
DB 146 GIOTLMGRLEDG---SPRTGQIFKQ 167

RESULT 14  
US-09-804-409A-16  
; Sequence 16, Application US/09804409A  
; Patent No. US20020155100A1  
; GENERAL INFORMATION:  
; APPLICANT: KIEFFER, TIMOTHY J.  
; APPLICANT: CHEUNG, ANTHONY I.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATED PROTEIN  
; TITLE OF INVENTION: EXPRESSION IN GUT  
; FILE REFERENCE: 029996/027 8721  
; CURRENT APPLICATION NUMBER: US/09/804,409A  
; CURRENT FILING DATE: 2001-03-12  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 16  
; LENGTH: 217  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-804-409A-16

Query Match 57.0%; Score 454; DB 10; Length 217;  
Best Local Similarity 69.7%; Pred. No. 2.8e-42;  
Matches 101; Conservative 7; Mismatches 21; Indels 16; Gaps 4;  
QY 2 FPTIPLSRFDNAMLRAHLHQLAFTDYOEFEEAIPKEQKYSFLQNPOTSLSPSESPT 61  
DB 27 FPTIPLSRFDNAMLRAHLHQLAFTDYOEFEEAIPKEQKYSFLQNPOTSLSPSESPT 86  
QY 62 PSNREETOQKSNLELLRLISLLIQSWLEPVLGTGPRFVNOHLCGS-----HLVE 111  
DB 87 PSNREETOQKSNLELLRLISLLIQSWLEPVLGTGPRFVNOHLCGS-----HLVE 145  
QY 112 ALYLVCG--ERGFFYTPKTRGIVEQ 134  
DB 146 GIOTLMGRLEDG---SPRTGQIFKQ 167

RESULT 15  
US-09-853-688-4  
; Sequence 4, Application US/09853688  
; Patent No. US20020081605A1  
; GENERAL INFORMATION:  
; APPLICANT: COOPER, DAVID N.  
; APPLICANT: PROCTER, ANNIE M.  
; APPLICANT: GREGORY, JOHN  
; APPLICANT: MILLAR, DAVID S.  
; TITLE OF INVENTION: METHOD FOR DETECTING GROWTH HORMONE VARIATIONS IN  
; FILE REFERENCE: WCH78  
; CURRENT APPLICATION NUMBER: US/09/853,688  
; CURRENT FILING DATE: 2001-05-14  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 217  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-853-688-4

Query Match 56.2%; Score 448; DB 9; Length 217;  
Best Local Similarity 69.0%; Pred. No. 1.3e-41;  
Matches 100; Conservative 7; Mismatches 22; Indels 16; Gaps 4;  
QY 2 FPTIPLSRFDNAMLRAHLHQLAFTDYOEFEEAIPKEQKYSFLQNPOTSLSPSESPT 61  
DB 27 FPTIPLSRFDNAMLRAHLHQLAFTDYOEFEEAIPKEQKYSFLQNPOTSLSPSESPT 86

Search completed: September 16, 2003, 12:52:26  
Job time : 47.2763 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2003, 12:36:15 : Search time 17.5097 Seconds  
(without alignments)  
823.845 Million cell updates/sec

Title: US-09-423-100-7

Perfect score: 797

Sequence: 1 MFPTIPLSRLFDNMLRAHR.....IVEQCSTCSISLYQLENYCN 150

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	461	57.8	217	1	STHU
2	460	57.7	217	1	SOMATOTROPIN 1 PRE
3	426.5	53.5	217	1	STHUV
4	426.5	53.5	256	1	SOMATOTROPIN 2 PRE
5	407.5	51.1	217	2	STHUV2
6	405	50.8	217	2	CHORIONIC SOMATOMA
7	396	49.7	212	2	SOMATOTROPIN - THE
8	396	49.7	212	2	CHORIONIC SOMATOMA
9	381	47.8	217	1	LCHUC
10	381	47.8	217	2	CHORIONAMMOTROPIN
11	359.5	45.1	215	2	CHORIONAMMOTROPIN
12	310.5	39.0	216	2	SOMATOTROPIN - GOL
13	307.5	38.6	190	2	SOMATOTROPIN - SEL
14	306.5	38.5	190	1	STHO
15	304.5	38.2	216	1	SOMATOTROPIN - HOR
16	302.5	38.0	216	1	STRT
17	302.5	38.0	216	2	SOMATOTROPIN PRECU
18	301.5	37.8	216	2	SOMATOTROPIN PRECU
19	301.5	37.8	216	1	STPG
20	301.5	37.8	216	2	SOMATOTROPIN PRECU
21	301.5	37.8	216	2	SOMATOTROPIN PRECU
22	299.5	37.6	216	2	SOMATOTROPIN PRECU
23	297.5	37.3	190	2	AJ7782
24	295.5	37.1	190	2	SOMATOTROPIN - ALP
25	289.5	36.3	217	1	STBO
26	289.5	36.3	217	1	STSH
27	289.5	36.3	217	1	STGT
28	289.5	36.3	217	2	S32582
29	276.5	34.9	216	2	JC1514

#### ALIGNMENTS

##### RESULT 1

STHU

Somatotropin 1 precursor [validated] - human

N:Alternate names: growth hormone 1; hGH-N; pituitary somatotropin

N:Contains: growth hormone 5K peptide; somatotropin 1, long form; somatotropin 1, shc

C:Species: Homo sapiens (man)

C>Date: 24-Apr-1984 #sequence revision 10-Feb-1995 #text change 08-Dec-2000

C:Accession: A93731; A32435; A93694; A94247; A90051; A93397; A93778; A91764; A90217;

R:DeNoto, F.M.; Moore, D.D.; Goodman, H.M.

Nucleic Acids Res. 9, 3719-3730, 1981

A:Title: Human growth hormone DNA sequence and mRNA structure: possible alternative s

A:Reference number: A93731; MUID:82014939; PMID:6269091

A:Accession: A93731

A:Molecule type: DNA

A:Residues: 1-217 <DEN>

A:Cross-references: GB:V00520

A>Note: the 20K short form somatoplin lacks residues 58-72 (32-45 in the active hor

R:Chen, E.Y.; Liao, Y.C.; Smith, T.; Barrera-Saldana, H.A.; Gelinas, R.E.; Seeburg,

Genomics 4, 479-497, 1989

A:Title: The human growth hormone cDNA sequence and nucleotide sequence, biology, and evolution

A:Reference number: A32435; MUID:1607277; PMID:2744760

A:Accession: A32435

A:Molecule type: DNA

A:Residues: 1-217 <CHE>

A:Cross-references: GB:V00571; MUID:1607148; PIRN:AAA52549.1; PIR:0184149

R:Roskar, W.; Kung'u, F.

Nucleic Acids Res. 7, 305-326, 1979

A:Title: Molecular cloning and nucleotide sequence of the human growth hormone struct

A:Reference number: A93694; MUID:80044477; PMID:386281

A:Accession: A93694

A:Molecule type: mRNA

A:Residues: 1-217 <ROS>

A:Cross-references: GB:V00519

A>Note: 35-pro was also found

R:Marital, J.A.; Halliwell, R.A.; Baxter, J.D.; Goodman, H.M.

Science 205, 602-607, 1979

A:Title: Human growth hormone: complementary DNA cloning and expression in bacteria.

A:Reference number: A94247; MUID:79203293; PMID:377496

A:Accession: A94247

A:Molecule type: mRNA

A:Residues: 1-217 <MAR>

R:Li, C.H.; Dixon, J.S.; Liu, Y.K.

Arch. Biochem. Biophys. 133, 70-91, 1969

A:Title: Human pituitary growth hormone. XIX. The primary structure of the hormone.

A:Reference number: A90048; MUID:69289202; PMID:5810834

A:Contents: annotation

R:Li, C.H.; Dixon, J.S.

Arch. Biochem. Biophys. 146, 233-236, 1971

A:Title: Human pituitary growth hormone. XXXII. The primary structure of the hormone.

A:Reference number: A90051; MUID:72143935; PMID:5144027

A:Accession: A90051

A:Molecule type: protein

insulin precursor  
insulin precursor  
somatotropin precu  
epidermal growth f  
insulin - sperm wh  
insulin - finback  
insulin - elephant  
insulin precursor  
insulin precursor  
insulin - hamster  
insulin precursor  
insulin - Egyptian  
somatotropin - gre  
insulin precursor  
insulin precursor  
insulin precursor

30 277.5 34.8 110 1 INRB  
31 277.5 34.8 110 2 B42179  
32 275.5 34.6 216 2 A60509  
33 275 34.5 96 2 PC7082  
34 273.5 34.3 51 1 L4HP  
35 273.5 34.3 51 1 L4HP  
36 273.5 34.3 51 1 INEL  
37 273.5 34.3 110 2 JQ0178  
38 272 34.1 110 2 A42179  
39 271.5 34.1 51 1 INHY  
40 270 33.9 110 1 IPHU  
41 268.5 33.7 51 1 INMSSP  
42 268.5 33.7 191 2 A60625  
43 267.5 33.6 51 2 A59151  
44 266.5 33.4 105 1 IPBO  
45 265.5 33.3 110 2 I48166

C:Comment: The gene for this hormone is transcribed only in somatotrophic cells of the pituitary gland.  
C:Gene: GDB:GH1  
A:Cross-references: GDB:119882; OMIM:139250  
A:Map position: 17q23.1-17q23.3  
A:Introns: 4/1; 57/3; 97/3; 152/3  
C:Superfamily: prolactin  
C:Keywords: alternative splicing; hormone; pituitary  
F:1-26/DNA: signal sequence #status predicted <SIG>  
F:27-217/Product: somatotropin 1, 19 kDa form #status experimental <SKP>  
F:27-69/Product: growth hormone 5K peptide #status experimental <SKP>  
F:27-57.73-217/Product: somatotropin 1, short form #status experimental <SOS>  
F:79-191,208-215/disulfide bonds: status experimental

Query Match 57.8% Score 461 DB 1 Length 217  
Best Local Similarity 70.3% ID No. 7.5e-38  
Matches 102 Conservative 7 Mismatches 20 Indels 16 Gaps 4

QY 2 FPTPLSLFDNMLRAHRLQLAFDTYQFEAYIPKEOKYSFLQNPTLSFSSES IPT 61  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 27 FPTPLSLFDNMLRAHRLQLAFDTYQFEAYIPKEOKYSFLQNPTLSFSSES IPT 86  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 62 PSNRRETOOKSNLELLRISLILQSLEPVOLGTGRFNQHLCGS-----HLVF 111  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 87 PSNRRETOOKSNLELLRISLILQSLEPVOLGTGRFNQHLCGS-----HLVF 145  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 112 ALYLVCG--ERGFYTPKTRCIYVQ 134  
:  
Db 146 GIOTLMGLIEDG---SPRTGTFKQ 167  
: :

RESULT 2  
I67410  
somatotropin - rhesus macaque  
N:Alternate names: growth hormone  
C:Species: Macaca mulatta (rhesus macaque)  
C>Date: 31-May-1996 Sequence Revision 31-May-1996 #text\_change 16-Jul-1999  
C:Accession: I67410; A05094  
R:Colos, T.G.; Durning, M.; Fisher, J.M.; Fowler, P.D.  
Endocrinology 133, 1744-1752, 1996  
A>Title: Cloning of four growth hormone/chorionic somatomotropin-related complement genes from the rhesus monkey  
A:Reference number: 153267; MUID:86129460; PMID:8404617  
A:Accession: I67410  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-217 <RES>  
R:Li, C.H.; Chung, D.; Lahm, H.W.; Stein, S.  
Arch. Biochem. Biophys. 245, 287-293, 1986  
A>Title: The primary structure of monkey pituitary growth hormone.  
A:Reference number: A05094; MUID:86129460; PMID:3080959  
A:Accession: A05094  
A:Molecule type: protein  
A:Residues: 27-99,'O','101-178,'D','180-217 <LIC>  
A>Note: the monkey species is not identified in the reference  
R:Raben, M.S.  
Science 125, 883-884, 1957  
A>Title: Preparation of growth hormone from pituitaries of man and monkey.  
A:Reference number: A44774  
A:Contents: annotation; Identification of source organism  
C:Superfamily: prolactin

Query Match 57.7% Score 460 DB 2 Length 217  
Best Local Similarity 98.9% ID No. 9.4e-38  
Matches 90 Conservative 1 Mismatches 1 Indels 0 Gaps 0

QY 2 FPTPLSLFDNMLRAHRLQLAFDTYQFEAYIPKEOKYSFLQNPTLSFSSES IPT 61  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 27 FPTPLSLFDNMLRAHRLQLAFDTYQFEAYIPKEOKYSFLQNPTLSFSSES IPT 86  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 62 PSNRRETOOKSNLELLRISLILQSLEPVOLGTGRFNQHLCGS-----HLVF 92  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

A:Residues: 27-94;96-217 <LIC>  
R:Niall, H.D.  
Nature New Biol. 230, 90-91, 1971  
A>Title: Revised primary structure for human growth hormone.  
A:Reference number: A93397; MUID:71139765; PMID:5279046  
A:Accession: A93397  
A:Molecule type: protein  
A:Residues: 27-51 <NIA>  
R:Niall, H.D.; Hogan, M.L.; Sauer, R.; Rosenblum, I.Y.; Greenwood, F.C.  
Proc. Natl. Acad. Sci. U.S.A. 68, 866-869, 1971  
A>Title: Sequences of pituitary and placental lactogenic and growth hormones: evolution  
A:Reference number: A93778; MUID:71153968; PMID:5279528  
A:Accession: A93778  
A:Molecule type: protein  
A:Residues: 119-120;157-159 <NI2>  
R:Niall, H.D.  
In Prolactin and Carcinogenesis, Proc. Fourth Tenovus Workshop Prolactin, Griffiths, K., Ed., Plenum Press, New York, 1978  
A>Title: The chemistry of the human lactogenic hormones.  
A:Reference number: A94427  
A:Contents: annotation; somatotropin revision  
R:Bewley, T.A.; Dixon, J.S.; Li, C.H.  
Int. J. Pept. Protein Res. 4, 281-287, 1972  
A>Title: Sequence comparison of human pituitary growth hormone, human chorionic somatotropic hormone, and porcine growth hormone.  
A:Reference number: A91764; MUID:73092028; PMID:4675454  
A:Accession: A91764  
A:Molecule type: protein  
A:Residues: 27-217 <BEW>  
R:Lewis, U.J.; Bonevald, L.F.; Lewis, L.J.  
Biochem. Biophys. Res. Commun. 92, 511-516, 1980  
A>Title: The 20,000-dalton variant of human growth hormone: location of the amino acid  
A:Reference number: A90217; MUID:80130196; PMID:7356479  
A:Contents: somatotropin, 20K short variant  
A:Accession: A90217  
A:Molecule type: protein  
A:Residues: 46-57;73-80 <LEW>  
R:Chapman, G.E.; Rogers, K.M.; Brittain, I.; Bradshaw, R.A.; Bates, O.J.; Turner, C.; Causton, B.  
J. Biol. Chem. 256, 2395-2401, 1981  
A>Title: The 20,000 molecular weight variant of human growth hormone. Preparation and  
A:Reference number: A92311; MUID:8117361; PM: 7462247  
A:Contents: somatotropin, 20K short variant  
A:Accession: A92311  
A:Molecule type: protein  
A:Residues: 27-57;73-79 <CHA>  
R:Singh, R.N.P.; Seavey, B.K.; Lewis, L.J.; Lewis, U.J.  
J. Protein Chem. 2, 425-436, 1983  
A>Title: Human growth hormone peptide 1-43: isolation from pituitary glands.  
A:Reference number: A61466  
A:Accession: A61466  
A:Molecule type: protein  
A:Residues: 27-69 <SIN>  
A>Note: growth hormone 5K peptide has insulin potentiating activity: its physiological  
R:Robson, V.M.J.; Rae, I.D.; NG, F.  
Biol. Chem. Hoppe-Seyler 371, 423-431, 1990  
A>Title: Identification of the aspartamide structure in a previously-reported peptide.  
A:Reference number: S09685; MUID:90334745; PMID:2378679  
A:Accession: S09685  
A:Molecule type: protein  
A:Residues: 27-34,'L','36-47 <ROB>  
R:de Vos, A.M.; Ultsch, M.; Kossiakoff, A.A.  
Science 255, 306-312, 1992  
A>Title: Human growth hormone and extracellular domain of its receptor: crystal structure  
A:Reference number: A41728; MUID:92196577; PMID:1549776  
A:Contents: annotation; X-ray crystallography, 2.8 angstroms  
A>Note: the structure of the complex with growth hormone receptor is described  
R:Gray, G.L.; Balbridge, J.S.; McKown, K.S.; Heyneker, H.L.; Chang, C.N.  
Gene 39, 247-254, 1985  
A>Title: Periplasmic production of correctly processed human growth hormone in Escherichia coli.  
A:Reference number: I41126; MUID:86137393; PMID:3912261  
A:Accession: I41126  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Residues: 1-26 <RES>  
A:Cross-references: GB:M14398; NID:g183158; PIDN:AAAS2554.1; PID:g183159



[illegible]

**RESULT 6**  
I67411  
somatotropin - rhesus macaque  
N.Alternate names: growth hormone  
C.Species: Macaca mulatta (rhesus macaque)  
C.Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 16-Jul-1999  
C.Accession: I67411  
R.Golos, T.G.; Durning, M.; Fisher, J.M.; Fowler, P.D.  
Endocrinology 133, 1744-1752, 1993  
A.Title: Cloning of four growth hormone/chorionic somatomammotropin-related complementary DNAs from rhesus macaque  
A.Reference number: I53267; MUID:94008724; PMID:8404617  
A.Accession: I67411  
A.Status: preliminary; translated from GB/EMBL/DBJ  
A.Molecule type: mRNA  
A.Residues: 1-217 <RES>  
A.Cross-references: GB:L16555; NID:g293116; PIDN:AAA20180.1; PID:g293117  
C.Superfamily: prolactin

Query Match 50.8%; Score 405; DB 2; Length 217;  
Best Local Similarity 67.7%; Pred. No. 2.4e-32;  
Matches 86; Conservative 9; Mismatches 18; Indels 14; Gaps 2;

QY 2 FTPIPLSRFDNAMLRAHRLHQLAFDTYOEFEEAYIPKEOKYSFLQNPTSLSFSESIPTP 61  
||||| ||| | : ||| ||||| : ||||| ||||| ||||| ||||| |||||  
DB 27 FTPIPLSMFNTAVFAVRAHLHLKLAFTDYPKEEAYIPKEOKYSERLPQTSLCFSESIPTP 86  
||||| ||| | : ||| ||||| : ||||| ||||| ||||| ||||| |||||  
QY 62 PSNRREETQSKNLELRISLLLIQSLEPVGCTGPRFNOLCGSHLVEA-----LY 114  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 87 PSNKETQSKNLELRISLLLIQSLEPVGF-----LRSVFANHLVHTNSFGLIYLY 139  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 115 LVCGERG 121  
| | | | |  
DB 140 LKKLEEG 146  
| | | | |

**RESULT 7**  
I67408  
Chorionic somatomammotropin-2 - rhesus macaque (fragment)  
C.Species: Macaca mulatta (rhesus macaque)  
C.Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 16-Jul-1999  
C.Accession: I67408  
R.Golos, T.G.; Durning, M.; Fisher, J.M.; Fowler, P.D.  
Endocrinology 133, 1744-1752, 1993  
A.Title: Cloning of four growth hormone/chorionic somatomammotropin-related complementary DNAs from rhesus macaque  
A.Reference number: I53267; MUID:94008724; PMID:8404617  
A.Accession: I67408  
A.Status: preliminary; translated from GB/EMBL/DBJ  
A.Molecule type: mRNA  
A.Residues: 1-212 <RES>  
A.Cross-references: GB:L16553; NID:g293110; PIDN:AAA18840.1; PID:g293111  
C.Superfamily: prolactin

Query Match 49.7%; Score 386; DB 2; Length 212;  
Best Local Similarity 82.2%; Pred. No. 1.8e-31;  
Matches 74; Conservative 11; Mismatches 5; Indels 0; Gaps 0;

QY 3 PTPILSRLFDNAMLRAHRLHQLAFDTYOEFEEAYIPKEOKYSFLQNPTSLSFSESIPTP 62  
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 23 PSVPILSRFDHAMIQARHLHQLAFDTYOEFEEAYIPKEKHSLMENPQASFCFADSIPTP 82  
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 63 SNREETOQSKNLELRISLLLIQSLEPVQ 92  
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 83 SNLEETOQSKNLELRISLLLIQSLEPVQ 112  
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

**RESULT 8**  
I53267  
Chorionic somatomammotropin-1 - rhesus macaque  
C.Species: Macaca mulatta (rhesus macaque)  
C.Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 16-Jul-1999  
C.Accession: I53267  
R.Golos, T.G.; Durning, M.; Fisher, J.M.; Fowler, P.D.  
Endocrinology 133, 1744-1752, 1993

```

QY      64 NREETQOKSNLELLRISLLLIQSWSLEPVQ 92
       I | | | | | | | | | | | | | | | | | | | |
DB      89 NMFEETQOKSNLELLRISLLLIQSWSLEPVR 117
       I | | | | | | | | | | | | | | | | | | | |

RESULT 10
E32435
Chorionamniototropin B precursor - human
N:Alternate names: chorionic somatomammotropin 2
C:Species: Homo sapiens (man)
C>Date: 29-Dec-1989 #sequence_revision 29-Dec-1989 #text_change 16-Jul-1999
C:Accession: 32435
R:Chen, E.Y.; Liao, Y.C.; Smith, D.H.; Bartera-Saldana, H.A.; Gelinas, R.E.; Seeburg,
Genomics 4, 479-497, 1989
A:title: The human growth hormone locus: nucleotide sequence, biology, and evolution
A:Reference number: A32435; MUID:89307277; PMID:2744760
A:Accession: E32435
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-217 <CHE>
A:CROSS-references: GB:J03071; NID:g183148; PIDN:AAA52553.1; PID:g183153
C:Genetics:
A:Gene: GDB:C5H2
A:CROSS-references: GDB:l19813; OMIM:118820
A:Map position: l7q22-l7q24
C:Superfamily: prolactin

Query Match          47.8%; Score 381; DB 2; Length 217,
Best local Similarity 82.08; Pred. No. 5,4e+30;
Matches 71; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

QY      4 TIPLSLRFNAMLRRAHLRHOLAFDTYOEEFAVIPKEOKYSFLONPOTSLSFSSES IPTPS 63
       I | | | | | | | | | | | | | | | | | | | |
DB      29 TVPLSLRFDAHLQAHRHAULADTYOEFETVYPKDQKYSPLHDSTSCFSDS IPTPS 88
       I | | | | | | | | | | | | | | | | | | | |

QY      64 NREETQOKSNLEILLRL LIQSWSLEPVQ 92
       I | | | | | | | | | | | | | | | | | | | |
DB      89 NMFEETQOKSNLELLRL LIQSWSLEPVR 117
       I | | | | | | | | | | | | | | | | | | | |

RESULT 11
A26449
Chorionamniototropin precursor (allele hCS-3) - human
C:Species: Homo sapiens (man)
C>Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 28-Jul-1995
C:Accession: A26449
R:Hirt, H.; Krimmelman, J.; Hirnbauch, M.J.; Chen, E.Y.; Seeburg, P.H.; Eberhardt, F.L.;
DNA 6, 59-70, 1987
A:title: The human growth hormone gene locus: structure, evolution, and allelic varia
A:Reference number: A26449; MUID:87161215; PMID:3030680
A:Accession: A26449
A:Molecule type: DNA
A:Residues: 1-215 <HIR>
C:Superfamily: prolactin
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-215/Product: chorionamniotropin, hCS-3 allele #status predicted <MAT>

Query Match          45.18; Score 359.5; DB 2; Length 215;
Best local Similarity 80.5%; Pred. No. 7e+28;
Matches 70; Conservative 1; Mismatches 8; Indels 1; Gaps 1;

QY      4 TIPLSLRFNAMLRRAHLRHOLAFDTYOEEFAVIPKEOKYSFLONPOTSLSFSSES IPTPS 63
       I | | | | | | | | | | | | | | | | | | | |
DB      29 TVPLSLRFDAHLQAHRHAULADTYOEFETVYPKDQKYSPLHDSTSCFSDS IPTPS 88
       I | | | | | | | | | | | | | | | | | | | |

QY      64 NREETQOKSNLELLRISLLLIQSWSLEP 90
       I | | | | | | | | | | | | | | | | | | | |
DB      89 NMFEETQOKSNLELLRL-LLLIQSWSLEP 114
       I | | | | | | | | | | | | | | | | | | | |

RESULT 12
B49159
somatotropin - golden hamster

```

A:Molecule type: protein  
A:Residues: 1-190 <ZAK>  
R:Zakin, M.M.; Poskusz, E.; Dellacha, J.M.; Paladini, A.C.; Santome, J.A.  
FEBS Lett. 34, 353-355, 1973  
A:Title: The amino acid sequence of equine growth hormone.  
A:Reference number: A91395; MUID:7:020362; PMID:4747849  
A:Accession: A91395  
A:Molecule type: protein  
A:Residues: 1-190 <ZAK>  
R:Zakin, M.M.; Poskusz, E.; Dellacha, J.M.; Paladini, A.C.; Santome, J.A.  
FEBS Lett. 25, 77-82, 1972  
A:Title: Amino acid sequences and the cysteine residues in equine growth hormone.  
A:Reference number: A91383  
A:Accession: A91383  
A:Molecule type: protein  
A:Residues: 42-69; 157-190 <ZA3>  
R:Oliver, L.; Hartree, A.S.  
Biochem. J. 109, 19-24, 1968  
A:Title: Amino acid sequences around the cysteine residues in horse growth hormone.  
A:Reference number: A90240; MUID:6:8468190; PMID:4876100  
A:Accession: A90240  
A:Molecule type: protein  
A:Residues: 176-190 <OI1>  
C:Superfamily: prolactin  
C:Keywords: hormone; pituitary  
F:52-163,180-188/Disulfide bonds: \*status experimental

Query Match 38.5%; Score 306.5; DB 1; Length 190;  
Best Local Similarity 65.2%; Ed. No. 9,7e-23;  
Matches 60; Conservative 14; Mismatches 17; Indels 1; Gaps 1;

QY 2 FPTIPLSRFDNAMLRAHRLHOIAFTDYOFFEAYIPKEQKYSFLONPOTSLSFSSES IPT 61  
II : III II III III III III III III III III III III III III III III  
DB 1 PSAMPSSLFANAVLRAQHILHQAADTYEFERKAYIPGEGRYS-IQNAQAACFSFIIPA 59  
II : III II III III III III III III III III III III III III III III  
QY 62 PSNPREETQOKSNLELLRIISLLLIQSWSLPVQL 93  
II : III III III III III III III III III III III III III III III III  
DB 60 PTCGDEAQQRSDMERLRFSLLLIQSWSLGPVQL 91  
II : III III III III III III III III III III III III III III III III

RESULT 15  
STMS  
Somatotropin precursor - mouse  
N:Alternate names: growth hormone  
C:Species: Mus musculus (house mouse)  
C>Date: 30-Sep-1987 #sequence revision 30-Sep-1987 #text change 28-May-1999  
C:Accession: R24911  
R:Linzer, D.F.H.; Tallantire, F.  
J. Biol. Chem. 260, 9574-9579, 1985  
A:Title: Nucleotide sequence of mouse prolactin and growth hormone mRNAs and expression  
A:Reference number: A92548; MUID:8:5263558; PMID:2991252  
A:Accession: R23911  
A:Molecule type: mRNA  
A:Residues: 1-216 <LIN>  
A:Cross-references: GB:X02891; GB:X03232; MID:951067; PIDN:CAA26650.1; PID:951068  
C:Superfamily: prolactin  
C:Keywords: anterior pituitary; growth factor; hormone  
F:1-26/Domain; signal sequence \*status predicted <SIG>  
F:27-216/Product; somatotropin \*status predicted <STN>  
F:78-189,206-214/Disulfide bonds: \*status predicted

Query Match 38.2%; Score 304.5; DB 1; Length 216;  
Best Local Similarity 64.8%; Pred. No. 1.8e-22;  
Matches 59; Conservative 14; Mismatches 17; Indels 1; Gaps 1;

QY 2 FPTIPLSRFDNAMLRAHRLHOIAFTDYOFFEAYIPKEQKYSFLONPOTSLSFSSES IPT 61  
II : III II III III III III III III III III III III III III III III  
DB 27 PSAMPSSLFANAVLRAQHILHQAADTYEFERKAYIPGEGRYS-IQNAQAACFSFIIPA 85  
II : III III III III III III III III III III III III III III III III  
QY 62 PSNPREETQOKSNLELLRIISLLLIQSWSLPVQL 92  
II : III III III III III III III III III III III III III III III III  
DB 86 PTCGEEAQORTOMELLRFSLLIQSWSLGPVQL 116  
II : III III III III III III III III III III III III III III III III

Search completed: September 16, 2003, 12:40:36  
Job time : 18.5097 secs

---

GenCore version 1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2003, 12:33:30 : Search time 11.6732 Seconds  
(without alignments)  
604.293 Million cell updates/sec

Title: us-09-423-100-7

Perfect score: 797

Sequence: 1 MFPTIPLSRLEFDNMLRAHR.....IVEQCCTSLCYQLENVCN 150

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	461	57.8	217	1 SOMA_HUMAN	P01241 homo sapien
2	461	57.8	217	1 SOMA_PANTR	P58756 pan troglod
3	460	57.7	217	1 SOMA_MACMU	P33093 macaca mula
4	437	54.8	217	1 SOMA_SAIIB	P58343 saimir bol
5	434.5	54.5	217	1 SOM2_PANTR	P58757 pan troglod
6	432	54.2	217	1 SOMA_CALJA	Q9amb3 callithrix
7	426.5	53.5	217	1 SOM2_HUMAN	P01242 homo sapien
8	399	50.1	217	1 SOM2_MACMU	Q07370 macaca mula
9	381	47.8	217	1 PLL_HUMAN	P01243 homo sapien
10	310.5	39.0	216	1 SOMA_MESAU	P37886 mesocricetu
11	307.5	38.6	190	1 SOMA_BALBO	P33092 balaeopter
12	306.5	38.5	216	1 SOMA_HORSE	P01245 equus cabal
13	306.5	38.5	217	1 SOMA_GALSE	Q9qkai galago sene
14	306.5	38.5	217	1 SOMA_NYCPY	Q9amb2 nycticebus
15	304.5	38.2	216	1 SOMA_MOUSE	P08880 mus musculu
16	302.5	38.0	216	1 SOMA_RABIT	P46407 oryctolagus
17	302.5	38.0	216	1 SOMA_RAT	P01244 rattus norv
18	301.5	37.8	190	1 SOMA_LOXAF	P20392 loxodonta a
19	301.5	37.8	216	1 SOMA_CANFA	P33711 canis fami
20	301.5	37.8	216	1 SOMA_FELCA	P46404 felis silve
21	301.5	37.8	216	1 SOMA_PIG	P01248 sus scrofa
22	299.5	37.6	216	1 SOMA_MUSVI	P19795 mustela vis
23	297.5	37.3	190	1 SOMA_LAMPA	P37885 lama guanac
24	295.5	37.1	190	1 SOMA_VULVU	P10766 vulpes vulp
25	291.5	36.6	215	1 SOMA_MONDO	Q9g160 monodelphis
26	291.5	36.6	215	1 SOMA_TRIVU	O62754 trichosurus
27	289.5	36.3	217	1 SOMA_BOVIN	P01246 bos taurus
28	289.5	36.3	217	1 SOMA_CEREL	P56437 cervus elap
29	289.5	36.3	217	1 SOMA_SHEEP	P01247 ovis aries
30	282.5	35.4	217	1 SOMA_BUBBU	O18938 bubalus bub
31	278.5	34.9	216	1 SOMA_MELGA	P22077 meleagris g
32	277.5	34.8	110	1 INS_CERAE	P30407 cercopithe
33	277.5	34.8	110	1 INS_RABIT	P01311 oryctolagu

#### RESULT 1

ID	SOMA_HUMAN	STANDARD	PRT	217 AA
AC	P01241: Q14405; Q16631; Q9HB21; Q9UMJ7; Q9UNL5;			
DI	21-JUL-1986 (Rel. 01, Created)			
DT	01-MAR-1992 (Rel. 21, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Somatotropin precursor (Growth hormone) (GH) (GH-N) (Pituitary growth hormone) (Growth hormone 1).			
GN	GH1			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A. (ISOFORM 1).			
RX	MEDLINE=80034477; PubMed=386281;			
RA	Roskam W., Rougeon F.;			
RT	"Molecular cloning and nucleotide sequence of the human growth hormone structural gene."			
RL	Nucleic Acids Res. 7:305-320(1979).			
RN	[2]			
RP	SEQUENCE FROM N.A. (ISOFORM 1).			
RX	MEDLINE=79203293; PubMed=377496;			
RA	Martial J.A., Halliwell R.A., Baxter J.D., Goodman H.M.;			
RT	"Human growth hormone: complementary DNA cloning and expression in bacteria."			
RL	Science 205:602-607(1979).			
RN	[3]			
RP	SEQUENCE FROM N.A. (ISOFORM 1), AND POSSIBLE ALTERNATIVE SPLICING.			
RX	MEDLINE=82014339; PubMed=6264691;			
RA	Denoto F.M., Moore D.D., Goodman H.M.;			
RT	"Human growth hormone DNA sequence and mRNA structure: possible alternative splicing."			
RL	Nucleic Acids Res. 9:3719-3730(1981).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=83182010; PubMed=7169009;			
RA	Seeburg P.H.;			
RT	"The human growth hormone gene family: nucleotide sequences show recent divergence and predict a new polypeptide hormone."			
RL	DNA 1:239-249(1982).			
RN	[5]			
RP	SEQUEN F FROM N.A.			
RX	MEDLINE=89307277; PubMed=274 760;			
RA	Chen E.Y., Liao Y.C., Smith J.H., Barrera-Saidana H.A.,			
RT	Gelinas R.E., Seeburg P.H.;			
RL	"The human growth hormone locus: nucleotide sequence, biology, and evolution."			
RN	[6]			
RP	SEQUENCE FROM N.A. (ISOFORM 3).			
RC	TISSUE=Pituitary;			
RA	Gu J., Huang Q.-H., Li N., Xu S.-H., Han Z.-G., Fu G., Chen Z.;			
RT	"A novel gene expressed in human pituitary."			
RL	Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.			

P08998 gallus galli  
Q9hwq3 struthio ca  
P01312 balaeopter  
P01316 eiephas max  
P30406 macaca fasc  
P55755 crocodylus  
P30410 pan troglod  
P01308 homo sapien  
P01324 acromys capi  
P34005 chelonias my  
P01317 bos taurus  
Q9ixi3 spermophilu

#### ALIGNMENTS

- RN [7]  
 RC SEQUENCE FROM N.A. (ISOFORM 4).  
 RX MEDLINE-20402571; PubMed-10931946;  
 RA Hu R.-M., Han Z.-G., Song H.-D., Peng Y.-D., Huang Q.-H., Ren S.-X.,  
 RA Gu Y.-J., Huang C.-H., Li Y.-B., Jiang C.-L., Fu G., Zhang Q.-H.,  
 RA Gu B.-W., Dai M., Mao Y.-F., Gao G.-F., Rong R., Ye M., Zhou J.,  
 RA Xu S.-H., Gu J., Shi J.-X., Jin W.-R., Zhang C.-K., Wu T.-M.,  
 RA Huang G.-Y., Chen Z., Chen M.-D., Chen J.-L.;  
 RT \*Gene expression profiling in the human hypothalamus-pituitary-adrenal  
 RT axis and full-length cDNA cloning.\*;  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:9543-9548(2000).  
 RN [18]  
 RN SEQUENCE OF 1-26 FROM N.A.  
 RX MEDLINE-86137393; PubMed-3912261;  
 RA Gray G.L., Baldrige J.S., McKeown K.S., Heyneker H.L., Chang C.N.;  
 RT \*Periplasmic production of correctly processed human growth hormone in  
 RT Escherichia coli: natural and bacterial signal sequences are  
 RT interchangeable.\*;  
 RL Gene 39:247-254(1985).  
 RN [19]  
 RN SEQUENCE OF 27-217.  
 RX MEDLINE-69289202; PubMed-5810834;  
 RA Li C.H., Dixon J.S., Liu W.-K.;  
 RT \*Human pituitary growth hormone. XIX. The primary structure of the  
 RT hormone.\*;  
 RL Arch. Biochem. Biophys. 133:70-91(1969).  
 RN [10]  
 RN SEQUENCE OF 27-217, AND REVISIONS.  
 RX MEDLINE-72143935; PubMed-5144027;  
 RA Li C.H., Dixon J.S.;  
 RT \*Human pituitary growth hormone. 32. The primary structure of the  
 RT hormone: revision.\*;  
 RL Arch. Biochem. Biophys. 146:233-236(1971).  
 RN [11]  
 RN REVISION.  
 RX MEDLINE-73092028; PubMed-4675454;  
 RA Bewley T.A., Dixon J.S., Li C.H.;  
 RT \*Sequence comparison of human pituitary growth hormone, human  
 RT chorionic somatomotropin, and ovine pituitary growth and  
 RT lactogenic hormones.\*;  
 RL Int. J. Pept. Protein Res. 4:281-287(1972).  
 RN [12]  
 RN SEQUENCE OF 27-61 AND 102-124.  
 RX MEDLINE-71139765; PubMed-5279046;  
 RA Niall H.D.;  
 RT \*Revised primary structure for human growth hormone.\*;  
 RL Nature New Biol. 230:90-91(1971).  
 RN [13]  
 RN REVISIONS TO 119-120 AND 157-159.  
 RX MEDLINE-71153968; PubMed-5279528;  
 RA Niall H.D., Hogan M.L., Sauer R., Rosenblum I.Y., Greenwood F.C.;  
 RT \*Sequences of pituitary and placental lactogenic and growth hormones:  
 RT evolution from a primordial peptide by gene reduplication.\*;  
 RL Proc. Natl. Acad. Sci. U.S.A. 68:866-869(1971).  
 RN [14]  
 RN REVISION.  
 RA Niall H.D.;  
 RT \*The chemistry of the human lactogenic hormones.\*;  
 RL Prolactin and carcinogenesis, proc. fourth tenovus workshop prolactin,  
 RL pp.13-20, Alpha Omega Alpha Press, Cardiff (1972).  
 RN [15]  
 RN SEQUENCE OF 27-79 (ISOFORM 2).  
 RX MEDLINE-81117161; PubMed-7462247;  
 RA Chapman G.E., Rogers K.M., Brittain T., Bradshaw R.A., Bates O.J.,  
 RA Turner C., Cary P.D., Crane-Robinson C.;  
 RT \*The 20,000 molecular weight variant of human growth hormone.  
 RT Preparation and some physical and chemical properties.\*;  
 RL J. Biol. Chem. 256:2395-2401(1981).  
 RN [16]  
 RN SEQUENCE OF 46-80 (ISOFORM 2).  
 RX MEDLINE-80130196; PubMed-7356479;  
 RA Lewis U.J., Bonewald L.F., Lewis L.J.;  
 RT \*The 20,000-dalton variant of human growth hormone: location of the  
 RT amino acid deletions.\*;  
 RL Biochem. Biophys. Res. Commun. 92:511-516(1980).  
 RN [17]  
 RN DEAMIDATION OF GLN-163 AND ASN-178.  
 RX MEDLINE-82052997; PubMed-7028740;  
 RA Lewis U.J., Singh R.N., Bonewald L.F., Seavey B.K.;  
 RT \*Altered proteolytic cleavage of human growth hormone as a result of  
 RT deamidation.\*;  
 RL J. Biol. Chem. 256:11645-11650(1981).  
 RN [18]  
 RN REVIEW.  
 RX MEDLINE-99321812; PubMed-10193484;  
 RA Baumann G.;  
 RT \*Growth hormone heterogeneity in human pituitary and plasma.\*;  
 RL Horm. Res. 51 Suppl. 1:2-6(1999).  
 RN [19]  
 RN 3D-STRUCTURE MODELING.  
 RX MEDLINE-88190073; PubMed-3447173;  
 RA Cohen F.E., Kuntz I.D.;  
 RT \*Prediction of the three-dimensional structure of human growth  
 RT hormone.\*;  
 RL Proteins 2:162-166(1987).  
 RN [20]  
 RN X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).  
 RX MEDLINE-92196577; PubMed-1549776;  
 RA de Vos A.M., Uitsch M., Kossiakoff A.A.;  
 RT \*Human growth hormone and extracellular domain of its receptor:  
 RT crystal structure of the complex.\*;  
 RL Science 255:306-312(1992).  
 RN [21]  
 RN X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS).  
 RX MEDLINE-95075462; PubMed-7984244;  
 RA Somers W., Uitsch M., de Vos A.M., Kossiakoff A.A.;  
 RT \*The X-ray structure of a growth hormone-prolactin receptor complex.\*;  
 RL Nature 371:478-481(1994).  
 RN [22]  
 RN X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
 RA Chantalat L., Chirgadze N.Y., Jones N., Korber F., Navaza J.,  
 RA Pavlovsk A.G., Wlodawer A.;  
 RT \*The crystal-structure of wild-type growth-hormone at 2.5-A  
 RT resolution.\*;  
 RL Protein Pept. Lett. 2:333-340(1995).  
 RN [23]  
 RN X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
 RX MEDLINE-97113024; PubMed-4943273;  
 RA Sundstroem M., Lundqvist T., Poedjin J., Giebel L.B., Milligan D.,  
 RA Norstedt G.;  
 RT \*Crystal structure of an antagonist mutant of human growth hormone,  
 RT G120R, in complex with its receptor at 2.9-A resolution.\*;  
 RL J. Biol. Chem. 271:32197-32203(1996).  
 RN [24]  
 RN VARIANT CYS-105.  
 RX MEDLINE-99318093; PubMed-10391209;  
 RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,  
 RA Shaw N., Lane C.R., Lim E.P., Kalyanaram N., Nemesh J., Ziaugra L.,  
 RA Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.O.,  
 RA Lander E.S.;  
 RT \*Characterization of single-nucleotide polymorphisms in coding regions  
 RT of human genes.\*;  
 RL Nat. Genet. 22:231-238(1999).  
 RN [25]  
 RN ERRATUM.  
 RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,  
 RA Shaw N., Lane C.R., Lim E.P., Kalyanaram N., Nemesh J., Ziaugra L.,  
 RA Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.O.,  
 RA Lander E.S.;  
 RL Nat. Genet. 23:373-373(1999).  
 CC -1- FUNCTION: Plays an important role in growth control. Its major  
 CC role in stimulating body growth is to stimulate the liver and  
 CC other tissues to secrete IGF-1. It stimulates both the  
 CC differentiation and proliferation of myoblasts. It also stimulates

Query Match 57.8%; Score 461; DB 1; Length 217;  
 Best Local Similarity 70.3%; Pred. No. 4.6e-38;  
 Matches 102; Conservative 7; Mismatches 20; Indels 16; Gaps 4;

QY 2 FPIPLSRFDNMLRAHRLHQLAFDTYQFEAYIPKEQKYSFLONPOTSLSFSESIPT 61  
 DQ 27 FPIPLSRFDNMLRAHRLHQLAFDTYQFEAYIPKEQKYSFLONPOTSLSFSESIPT 86  
 QY 62 PSNRETOQKSNLELLRISLLLIQSWLEPVQLGTGPRFVNOHLGCS-----HLVE 111  
 DQ 87 PSNRETOQKSNLELLRISLLLIQSWLEPVQF--LRSVFANSFLVYGASDSNYVDLLKDL 145  
 QY 112 ALYLVCG--ERGFFYTPKTRGIVEQ 134  
 DQ 146 GIOTLMGRLEDG---SPRTGQIFKQ 167

RESULT 2  
 SOMA\_PANTR STANDARD; PRT; 217 AA.  
 AC P58756;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Somatotropin precursor (Growth hormone) (GH) (GH-N) (Pituitary growth hormone) (Growth hormone 1).  
 GN GH1.  
 OS Macaca mulatta (Rhesus macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
 OX NCBI\_TaxID=9598;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Revol A., Esquivel D., Santiago D., Barrera-Saldana H.;  
 RT "Independent duplication of the growth hormone gene in three Anthropoid lineages."  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Plays an important role in growth control. Its major role in stimulating body growth is to stimulate the liver and other tissues to secrete IGF-1. It stimulates both the differentiation and proliferation of myoblasts. It also stimulates amino acid uptake and protein synthesis in muscle and other tissues (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: AF374232; AAL72284.1;  
 CC InterPro: IPR001400; Somatotropin.  
 CC Pfam: PF00103; hormone.1.  
 CC PRINTS: PR00836; SOMATOTROPIN.  
 CC PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
 CC PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
 CC Hormone; Pituitary; Signal.  
 CC SIGNAL 1 26 BY SIMILARITY.  
 CC CHAIN 27 217 SOMATOTROPIN.  
 CC DISULFID 79 191 BY SIMILARITY.  
 CC DISULFID 208 215 BY SIMILARITY.  
 CC CONFLICT 100 100 E -> Q (IN REF. 2).  
 CC CONFLICT 179 179 N -> D (IN REF. 2).  
 CC SEQUENCE 217 AA; 24843 MW; FEA295DE0518674 CRC64;

Query Match 57.8%; Score 461; DB 1; Length 217;  
 Best Local Similarity 70.3%; Pred. No. 4.6e-38;  
 Matches 102; Conservative 7; Mismatches 20; Indels 16; Gaps 4;

QY 2 FPIPLSRFDNMLRAHRLHQLAFDTYQFEAYIPKEQKYSFLONPOTSLSFSESIPT 61  
 DQ 27 FPIPLSRFDNMLRAHRLHQLAFDTYQFEAYIPKEQKYSFLONPOTSLSFSESIPT 86  
 QY 62 PSNRETOQKSNLELLRISLLLIQSWLEPVQLGTGPRFVNOHLGCS-----HLVE 111  
 DQ 87 PSNRETOQKSNLELLRISLLLIQSWLEPVQF--LRSVFANSFLVYGASDSNYVDLLKDL 145  
 QY 112 ALYLVCG--ERGFFYTPKTRGIVEQ 134  
 DQ 146 GIOTLMGRLEDG---SPRTGQIFKQ 167

Db 27 FPIPLSRFDNMLRAHRLHQLAFDTYQFEAYIPKEQKYSFLONPOTSLSFSESIPT 86  
 QY 62 PSNRETOQKSNLELLRISLLLIQSWLEPVQLGTGPRFVNOHLGCS-----HLVE 111  
 Db 87 PSNRETOQKSNLELLRISLLLIQSWLEPVQF--LRSVFANSFLVYGASDSNYVDLLKDL 145  
 QY 112 ALYLVCG--ERGFFYTPKTRGIVEQ 134  
 Db 146 GIOTLMGRLEDG---SPRTGQIFKQ 167

RESULT 3  
 SOMA\_MACMU STANDARD; PKT; 217 AA.  
 AC P33093;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Somatotropin precursor (Growth hormone) (GH) (GH-N) (Pituitary growth hormone) (Growth hormone 1).  
 GN GH1.  
 OS Macaca mulatta (Rhesus macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae; Macaca.  
 OX NCBI\_TaxID=9544;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=94008724; PubMed=8404.17;  
 RT "Cloning of four growth hormone/chorionic somatomammotropin-related complementary deoxyribonucleic acids differentially expressed during pregnancy in the rhesus monkey placenta."  
 RL Endocrinology 133:1744-1752 (1993).  
 RN [2]  
 RP SEQUENCE OF 27-217.  
 RA MEDLINE=86129460; PubMed=3080959;  
 RA Li C.H., Chung D., Lahn H.W., Stein S.;  
 RT "The primary structure of monkey pituitary growth hormone."  
 RL Arch. Biochem. Biophys. 245:287-291 (1986).  
 CC -1- FUNCTION: Plays an important role in growth control. Its major role in stimulating body growth is to stimulate the liver and other tissues to secrete IGF-1. It stimulates both the differentiation and proliferation of myoblasts. It also stimulates amino acid uptake and protein synthesis in muscle and other tissues.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: L16556; AAA18842.1;  
 CC PIR: I67410; I67410.  
 CC HSP: P01241; IAXI.  
 CC InterPro: IPR001400; Somatotropin.  
 CC Pfam: PF00103; hormone.1.  
 CC PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
 CC PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
 CC Hormone; Pituitary; Signal.  
 CC SIGNAL 1 26 BY SIMILARITY.  
 CC CHAIN 27 217 SOMATOTROPIN.  
 CC DISULFID 79 191 BY SIMILARITY.  
 CC DISULFID 208 215 BY SIMILARITY.  
 CC CONFLICT 100 100 E -> Q (IN REF. 2).  
 CC CONFLICT 179 179 N -> D (IN REF. 2).  
 CC SEQUENCE 217 AA; 24913 MW; 2C5180341EEC46D0 CRC64;

Query Match 57.7%; Score 460; DB 1; Length 217;  
Best Local Similarity 98.9%; Pred. No. 5.8e-38;  
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 FPTPLSLFONAMLRHRLHQLAFDTYQEEAYIPKEQKYSFLONPQTSLSFSESPT 61  
DB 27 FPTPLSLFONAMLRHRLHQLAFDTYQEEAYIPKEQKYSFLONPQTSLSFSESPT 86  
QY 62 PSNRETOOKSNLELLRLISLLLIQSWLEPVQ 92  
DB 87 PSNRETOOKSNLELLRLISLLLIQSWLEPVQ 117

RESULT 4  
SOMA\_SAIBB STANDARD: PRT: 217 AA.  
AC P58343;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Somatotropin precursor (Growth hormone).  
GN GH1.  
OS Salmirol boliviensis boliviensis (Bolivian squirrel monkey).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae; Saimiri.  
OX NCBI\_TaxID=39432;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21265430; PubMed=11371582;  
RA Liu J.C., Makova K.D., Adkins R.M., Gibson S., Li W.H.;  
RT "Episodic evolution of growth hormone in primates and emergence of the  
species specificity of human growth hormone receptor";  
RL Mol. Biol. Evol. 18:945-953(2001).  
CC -!- FUNCTION: Plays an important role in growth control. Its major  
role in stimulating body growth is to stimulate the liver and  
other tissues to secrete IGF-1. It stimulates both the  
differentiation and proliferation of myoblasts. It also stimulates  
amino acid uptake and protein synthesis in muscle and other  
tissues (By similarity).  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: AF330600; AAK62287.1;  
InterPro: IPR001400; Somatotropin.  
DR Pfam: PF00103; hormone. 1.  
DR PRINTS: P00836; SOMATOTROPIN.  
DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
KW Hormone; Pituitary; Signal.  
FT SIGNAL 1 26 BY SIMILARITY.  
FT CHAIN 27 217 SOMATOTROPIN.  
FT DISULFID 79 191 BY SIMILARITY.  
FT DISULFID 208 215 BY SIMILARITY.  
SQ SEQUENCE 217 AA; 24864 MW; 9515289992C529F7 CRC64;

Query Match 54.8%; Score 437; DB 1; Length 217;  
Best Local Similarity 91.3%; Pred. No. 1e-35;  
Matches 84; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 FPTPLSLFONAMLRHRLHQLAFDTYQEEAYIPKEQKYSFLONPQTSLSFSESPT 61  
DB 27 FPTPLSLFONAMLRHRLHQLAFDTYQEEAYIPKEQKYSFLONPQTSLSFSESPT 86  
QY 62 PSNRETOOKSNLELLRLISLLLIQSWLEPVQ 93

DB 87 PASKKETQOKSNLELLRLISLLLIQSWLEPVQ 118

RESULT 5  
SOMA\_PANTR STANDARD: PRT: 217 AA.  
AC P58757;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Growth hormone variant precursor (GH-V) (Placenta-specific growth  
hormone) (Growth hormone 2).  
GN GH2.  
OS Pan troglodytes (Chimpanzee).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Pan.  
OX NCBI\_TaxID=9598;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Revol A., Esquivel D., Sanriago D., Bartera-Saidana H.;  
RT "Independent duplication of the growth hormone gene in three  
Anthropoid lineages";  
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Plays an important role in growth control. Its major  
role in stimulating body growth is to stimulate the liver and  
other tissues to secrete IGF-1. It stimulates both the  
differentiation and proliferation of myoblasts. It also stimulates  
amino acid uptake and protein synthesis in muscle and other  
tissues.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- TISSUE SPECIFICITY: Expressed in the placenta.  
CC -!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: AF374233; AAL72285.1;  
InterPro: IPR001400; Somatotropin.  
DR Pfam: PF00103; hormone. 1.  
DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
KW Hormone; Placenta; Signal; Glycoprotein.  
FT SIGNAL 1 26 BY SIMILARITY.  
FT CHAIN 27 217 GROWTH HORMONE VARIANT.  
FT DISULFID 79 191 BY SIMILARITY.  
FT DISULFID 208 215 BY SIMILARITY.  
SQ SEQUENCE 217 AA; 24990 MW; 1592A429075677DE CRC64;

Query Match 54.5%; Score 434.5; DB 1; Length 217;  
Best Local Similarity 78.9%; Pred. No. 1.8e-35;  
Matches 90; Conservative 4; Mismatches 9; Indels 11; Gaps 1;

QY 2 FPTPLSLFONAMLRHRLHQLAFDTYQEEAYIPKEQKYSFLONPQTSLSFSESPT 61  
DB 27 FPTPLSLFONAMLRHRLHQLAFDTYQEEAYIPKEQKYSFLONPQTSLSFSESPT 86  
QY 62 PSNRETOOKSNLELLRLISLLLIQSWLEPVQ 104  
DB 87 PSNRVKTQOKSNLELLRLISLLLIQSWLEPVQ 140

RESULT 6  
SOMA\_CALJA STANDARD: PRT: 217 AA.  
ID SOMA\_CALJA  
AC Q9GMB3;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)

28-FEB-2003 (Rel. 41, Last annotation update)  
Somatotropin precursor (Growth hormone).  
GHI.  
Callitrix jacchus (Common marmoset).  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;  
Callitrix.  
NCBI\_TaxID=9483;  
[1]  
SEQUENCE FROM N.A.  
Wallis O.C., Wallis M.;  
"Cloning and characterisation of a putative growth hormone encoding  
gene from the marmoset (Callitrix jacchus).";  
Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
-!- FUNCTION: Plays an important role in growth control. Its major  
role in stimulating body growth is to stimulate the liver and  
other tissues to secrete IGF-1. It stimulates both the  
differentiation and proliferation of myoblasts. It also stimulates  
amino acid uptake and protein synthesis in muscle and other  
tissues (By similarity).  
-!- SUBCELLULAR LOCATION: Secreted.  
-!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
EMBL; AJ297563; CAC03481.1; -  
HSSP: P01241; 1A22.  
InterPro: IPR001400; Somatotropin.  
Pfam: PF00103; hormone\_1  
PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
Hormone; pituitary; Signal.  
SIGNAL 1 26 BY SIMILARITY.  
CHAIN 27 217 SOMATOTROPIN.  
DISULFID 79 191 BY SIMILARITY.  
DISULFID 208 215 BY SIMILARITY.  
SEQUENCE 217 AA; 24959 MW; E102151A12CF6192 CRC64;  
Query Match 54.28; Score 432; DB i; Length 217;  
Best Local Similarity 91.28; Pred. No. 3.28-35;  
Matches 83; Conservative 5; Mismatches 3; Indels 0; Gaps 0;  
QY 2 FPTPLSLRFONAMLAHRLHQLAFTDYQPEEAYIPKEQKYSFLQNPQTSLSISEPT 61  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
27 FPTPLSLRLDNAMLAHRLHQLAFTDYQPEEAYIPKEQKYSFLQNPQTSLSISEPT 86  
QY 62 PSNREETQOKSNLELLRLSLLIOSWLEPVQ 92  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
DB 87 PASKRETOOKSNLELLRLSLLIOSWLEPVQ 117  
RESULT 7  
RESOM2\_HUMAN  
ID SOM2\_HUMAN STANDARD; PRT; 217 AA.  
AC P01242; P09587;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Growth hormone variant precursor (GH-V) (Placenta-specific growth  
hormone) (Growth hormone 2).  
OS Homo sapiens (Human).  
GN GH2.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RA Seeburg P.H.;  
"The human growth hormone gene family: nucleotide sequences show  
recent divergence and predict a new polypeptide hormone.";  
DNA 1:239-249(1982).  
[2]  
SEQUENCE FROM N.A. (ISOFORM 1 AND 2).  
RA Cooke N.E., Ray J., Emery J.J., Liebhauer S.A.;  
MEDLINE=88243769; PubMed=33 057;  
"Two distinct species of human growth hormone-variant mRNA in the  
human placenta predict the expression of novel growth hormone  
proteins.";  
J. Biol. Chem. 263:9001-9006(1988).  
[3]  
SEQUENCE FROM N.A. (ISOFORM 1).  
MEDLINE=89024984; PubMed=2460050;  
RA Igout A., Scippo M.L., Frankeu F., Hennen G.;  
"Cloning and nucleotide sequence of placental hGH-V cDNA".;  
Arch. Int. Physiol. Biochim. 96:63-67(1988).  
[4]  
SEQUENCE FROM N.A.  
MEDLINE=89307277; PubMed=27 4760;  
RA Chen E.Y., Liao Y.C., Smith L.H., Barrera-Saldana H.A.,  
Gallinas R.E., Seeburg P.H.;  
"The human growth hormone locus: nucleotide sequence, biology, and  
evolution".;  
Genomics 4:479-497(1989).  
[5]  
SEQUENCE FROM N.A.  
TISSUE=Placenta;  
MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
Altschul S.F., Zeeberg B., Luetow K.H., Schaefer C.F., Bhat N.K.,  
Hopkins R.F., Jordan H., Moore J., Max S.L., Wang J., Wieh F.,  
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
Buetterfield Y.S.N., Krzywinski M.I., Skalski U., Smalley D.F.,  
Schnerf A., Schein J.E., Jones S.J.M., Marra M.A.;  
"Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences".;  
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
[6]  
REVIEW  
MEDLINE=93321812; PubMed=10393484;  
RA Baumann G.;  
"Growth hormone heterogeneity in human pituitary and plasma".;  
Horm. Res. 51 Suppl. 1:2-6(1999).  
-!- FUNCTION: Plays an important role in growth control. Its major  
role in stimulating body growth is to stimulate the liver and  
other tissues to secrete IGF-1. It stimulates both the  
differentiation and proliferation of myoblasts. It also stimulates  
amino acid uptake and protein synthesis in muscle and other  
tissues.  
-!- SUBUNIT: Monomer, dimer, trimer, tetramer and pentamer, disulfide-  
linked or non-covalently associated, in homopolymetric and  
heteropolymetric combinations. Can also form a complex either with  
GHBP or with the alpha2-macroglobulin complex.  
-!- SUBCELLULAR LOCATION: Secreted.  
-!- ALTERNATIVE PRODUCTS:  
Event-Alternative splicing; Named isoforms=2;  
Name=1; Synonyms=GH-V1;  
IsoId=P01242.1; Sequence=Displayed;  
Name=2; Synonyms=GH-V2;  
SEQUENCE FROM N.A. (ISOFORM 1).  
RA Seeburg P.H.;  
"The human growth hormone gene family: nucleotide sequences show  
recent divergence and predict a new polypeptide hormone".;  
DNA 1:239-249(1982).  
[2]  
SEQUENCE FROM N.A. (ISOFORM 1 AND 2).  
RA Cooke N.E., Ray J., Emery J.J., Liebhauer S.A.;  
MEDLINE=88243769; PubMed=33 057;  
"Two distinct species of human growth hormone-variant mRNA in the  
human placenta predict the expression of novel growth hormone  
proteins".;  
J. Biol. Chem. 263:9001-9006(1988).  
[3]  
SEQUENCE FROM N.A. (ISOFORM 1).  
MEDLINE=89024984; PubMed=2460050;  
RA Igout A., Scippo M.L., Frankeu F., Hennen G.;  
"Cloning and nucleotide sequence of placental hGH-V cDNA".;  
Arch. Int. Physiol. Biochim. 96:63-67(1988).  
[4]  
SEQUENCE FROM N.A.  
MEDLINE=89307277; PubMed=27 4760;  
RA Chen E.Y., Liao Y.C., Smith L.H., Barrera-Saldana H.A.,  
Gallinas R.E., Seeburg P.H.;  
"The human growth hormone locus: nucleotide sequence, biology, and  
evolution".;  
Genomics 4:479-497(1989).  
[5]  
SEQUENCE FROM N.A.  
TISSUE=Placenta;  
MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
Altschul S.F., Zeeberg B., Luetow K.H., Schaefer C.F., Bhat N.K.,  
Hopkins R.F., Jordan H., Moore J., Max S.L., Wang J., Wieh F.,  
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
Buetterfield Y.S.N., Krzywinski M.I., Skalski U., Smalley D.F.,  
Schnerf A., Schein J.E., Jones S.J.M., Marra M.A.;  
"Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences".;  
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
[6]  
REVIEW  
MEDLINE=93321812; PubMed=10393484;  
RA Baumann G.;  
"Growth hormone heterogeneity in human pituitary and plasma".;  
Horm. Res. 51 Suppl. 1:2-6(1999).  
-!- FUNCTION: Plays an important role in growth control. Its major  
role in stimulating body growth is to stimulate the liver and  
other tissues to secrete IGF-1. It stimulates both the  
differentiation and proliferation of myoblasts. It also stimulates  
amino acid uptake and protein synthesis in muscle and other  
tissues.  
-!- SUBUNIT: Monomer, dimer, trimer, tetramer and pentamer, disulfide-  
linked or non-covalently associated, in homopolymetric and  
heteropolymetric combinations. Can also form a complex either with  
GHBP or with the alpha2-macroglobulin complex.  
-!- SUBCELLULAR LOCATION: Secreted.  
-!- ALTERNATIVE PRODUCTS:  
Event-Alternative splicing; Named isoforms=2;  
Name=1; Synonyms=GH-V1;  
IsoId=P01242.1; Sequence=Displayed;  
Name=2; Synonyms=GH-V2;  
SEQUENCE FROM N.A. (ISOFORM 1).  
RA Seeburg P.H.;  
"The human growth hormone gene family: nucleotide sequences show  
recent divergence and predict a new polypeptide hormone".;  
DNA 1:239-249(1982).  
[2]  
SEQUENCE FROM N.A. (ISOFORM 1 AND 2).  
RA Cooke N.E., Ray J., Emery J.J., Liebhauer S.A.;  
MEDLINE=88243769; PubMed=33 057;  
"Two distinct species of human growth hormone-variant mRNA in the  
human placenta predict the expression of novel growth hormone  
proteins".;  
J. Biol. Chem. 263:9001-9006(1988).  
[3]  
SEQUENCE FROM N

```

CC      IsoId-P01242-2: Sequence-VSP_006203:
CC      Note-No experimental confirmation available;
CC      -1- TISSUE SPECIFICITY: Expressed in the placenta.
CC      -1- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: K00470; AAA98619.1;
CC      EMBL: J03756; AAB59547.1;
CC      EMBL: J03756; AAB59548.1;
CC      EMBL: M38451; AAA35891.1;
CC      EMBL: J03071; AAA52552.1;
CC      EMBL: BC020760; AAH20760.1;
CC      PIR: A28072; STH0V2.
CC      PIR: D32435; STH0V.
CC      HSSP: P01241; 1A22.
CC      Genew: HGNC:4262; GH2.
CC      MW: 139240.
CC      GO: GO:0005180; F:peptide hormone; TAS.
CC      InterPro: IPR001400; Somatotropin.
CC      Pfam: PF00103; hormone; 1.
CC      PROSITE: PS00266; SOMATOTROPIN_1; 1.
CC      PROSITE: PS00338; SOMATOTROPIN_2; 1.
CC      Hormone; Placenta; Signal: Glycoprotein; Alternative splicing:
CC      Polymorphism.
CC      SIGNAL      1 26
CC      CHAIN      27 217 GROWTH HORMONE VARIANT.
CC      DISULFID   79 191 BY SIMILARITY.
CC      DISULFID   208 215 N-LINKED (GLCNAC...) (POTENTIAL).
CC      CARBOHYD   166 166 RLEDSPTGQIFNQSYSKFDTKSHNDAILKNTGLLYCFR
CC      VARSPLIC   153 217 KMDKYETFLRVQSVSGVEGCGF -> VRVAPGIPNPAP
CC      LASRDWGEKHC: LFSSQALTOGSPSYSSFLVNPVGLSLQ
CC      PGEGGKWNK: REQCPSPALLLFLFAEAGRWQPPDWA
CC      DIOSVLOOV ( ) Isoform 2).
CC      FT VARIANT      90 90 R -> W (IN dbSNP:5389).
CC      FT CONFLICT    109 109 /FTIG-VAR_014591.
CC      FT SEQUENCE    217 AA: 24999 MW: 789324698R22F96 Cht64.
CC      -----
CC      Query Match      53.5%; Score 426.5; DB 1; Length 217;
CC      Best Local Similarity 78.1%; Pred. No. 1.1e-34;
CC      Matches 89; Conservative 4; Mismatches 10; Indels 11; Gaps 1;
CC      -----
CC      2 FPTIPLSRFDNMLRAHLHQLAFDTYQEFEEVYIPKEQYSLQNPQTSLSFSESPT 61
CC      27 FPTIPLSRFDNMLRAHLHQLAFDTYQEFEEVYIPKEQYSLQNPQTSLSFSESPT 86
CC      -----
CC      62 PSNRRETOOKSNLELLRISLLTIQSWLEPVLQGTGRFVNOHLCGSHLVEA-----LY 114
CC      87 PSNRVATQOKSNLELLRISLLTIQSWLEPVLQGTGRFVNOHLCGSHLVEA-----LY 139
CC      -----
CC      RESULT 8
CC      SOM2_MACMU STANDARD: PRI: 217 AA.
CC      ID SOM2_MACMU
CC      AC Q07370; Q28494.
CC      DT 01-NOV-1997 (Rel. 35, Created)
CC      DT 01-NOV-1997 (Rel. 35, Last sequence update)
CC      DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC      DE Growth hormone variant precursor (GH-V) (Placenta-specific growth
CC      DE hormone) (Growth hormone 2).
CC      GN GH2.
CC      OS Macaca mulatta (Rhesus macaque).
CC      OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC      OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;

```

```

CC      Cercopitheidae; Macaca.
CC      NCBI_TaxID-9544;
CC      [1]
CC      RP SEQUENCE FROM N.A.
CC      RA Golos T.G.;
CC      Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
CC      [2]
CC      RP SEQUENCE FROM N.A.
CC      MEDLINE-94008724; PubMed-8404617;
CC      RA Golos T.G.; Durning M.; Fisher J.M.; Fowler P.D.;
CC      *Cloning of four growth hormone/chorionic somatomammotropin-related
CC      complementary deoxyribonucleic acids differentially expressed during
CC      pregnancy in the rhesus monkey placenta.*;
CC      Endocrinology 133:1744-1752(1993).
CC      -1- FUNCTION: Plays an important role in growth control. Its major
CC      role in stimulating body growth is to stimulate the liver and
CC      other tissues to secrete IGF-1. It stimulates both the
CC      differentiation and proliferation of myoblasts. It also stimulates
CC      amino acid uptake and protein synthesis in muscle and other
CC      tissues.
CC      -1- SUBCELLULAR LOCATION: Secreted (By similarity).
CC      -1- TISSUE SPECIFICITY: Expressed in the placenta.
CC      -1- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: U02293; AAA03391.1;
CC      EMBL: L16555; AAA20180.1;
CC      PIR: I67411; I67411.
CC      HSSP: P01241; 1HGU.
CC      InterPro: IPR001400; Somatotropin.
CC      Pfam: PF00103; hormone; 1.
CC      PRINTS: PR00836; SOMATOTROPIN.
CC      PROSITE: PS00266; SOMATOTROPIN_1; 1.
CC      PROSITE: PS00338; SOMATOTROPIN_2; 1.
CC      Hormone; Placenta; Signal: Glycoprotein.
CC      SIGNAL      1 26 BY SIMILARITY.
CC      CHAIN      27 217 GROWTH HORMONE VARIANT.
CC      DISULFID   79 191 BY SIMILARITY.
CC      DISULFID   208 215 BY SIMILARITY.
CC      CONFLICT    57 57 L -> F (IN REF. 2).
CC      CONFLICT    152 152 E -> G (IN REF. 2).
CC      FT SEQUENCE    217 AA: 25221 MW: 8DB116CBC24EA090 Cht64;
CC      -----
CC      Query Match      50.1%; Score 399; DB 1; Length 217;
CC      Best Local Similarity 66.9%; Pred. No. 5.4e-32;
CC      Matches 85; Conservative 9; Mismatches 19; Indels 14; Gaps 2;
CC      -----
CC      2 FPTIPLSRFDNMLRAHLHQLAFDTYQEFEEVYIPKEQYSLQNPQTSLSFSESPT 61
CC      27 FPTIPLSRFDNMLRAHLHQLAFDTYQEFEEVYIPKEQYSLQNPQTSLSFSESPT 86
CC      -----
CC      62 PSNRRETOOKSNLELLRISLLTIQSWLEPVLQGTGRFVNOHLCGSHLVEA-----LY 114
CC      87 PSNRRETOOKSNLELLRISLLTIQSWLEPVLQGTGRFVNOHLCGSHLVEA-----LY 139
CC      -----
CC      RESULT 9
CC      PLL_HUMAN
CC      ID PLL_HUMAN
CC      AC P01243;
CC      DT 21-JUL-1986 (Rel. 01, Created)

```

DT 01-APR-1988 (Rel. 07, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Lactogen precursor (Choriomammotropin) (Chorionic somatomammotropin).  
 GN CSH1 AND CSH2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RN SEQUENCE FROM N.A. (GENE CSH1).  
 RP MEDLINE-85030426; PubMed-6208192;  
 RX Selby M.J., Barta A., Baxter J.D., Bell G.I., Eberhardt N.L.;  
 RA "Analysis of a major human chorionic somatomammotropin gene. Evidence  
 RT for two functional promoter elements";  
 RL J. Biol. Chem. 259:13131-13138(1984).  
 RN [2]  
 RN SEQUENCE FROM N.A. (GENE CSH2).  
 RP MEDLINE-87161235; PubMed-3030680;  
 RX Hirt H., Kimmelman J., Birnbaum M.J., Chen E.Y., Seeburg P.H.,  
 RA Eberhardt N.L., Barta A.;  
 RT "The human growth hormone gene locus: structure, evolution, and  
 RT allelic variations";  
 RL DNA 6:59-70(1987).  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE-83160916; PubMed-6300056;  
 RX Barrera-Saldana H.A., Seeburg P.H., Saunders G.F.;  
 RA "Two structurally different genes produce the same secreted human  
 RT placental lactogen hormone";  
 RL J. Biol. Chem. 258:3787-3793(1983).  
 RN [4]  
 RN SEQUENCE FROM N.A. (GENES CSH1 AND CSH2).  
 RP MEDLINE-89302777; PubMed-2744760;  
 RX Chen E.Y., Liao Y.C., Smith D.H., Barrera-Saldana H.A., Gelinas R.E.,  
 RA Seeburg P.H.;  
 RT "The human growth hormone locus: nucleotide sequence, biology, and  
 RT evolution";  
 RL Genomics 4:479-497(1989).  
 RN [5]  
 RN SEQUENCE.  
 RP MEDLINE-83182010; PubMed-7169009;  
 RX Seeburg P.H.;  
 RA "The human growth hormone gene family: nucleotide sequences show  
 RT recent divergence and predict a new polypeptide hormone";  
 RL DNA 1:239-249(1982).  
 RN [6]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=Placenta, and Uterus;  
 MEDLINE-22388257; PubMed-12477432;  
 RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,  
 RA Sapletenko M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Murry D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting J., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [7]  
 RN SEQUENCE (F 50-217 FROM N.A.  
 RX MEDLINE-78071761; PubMed-593368;  
 RA Shine J., Seeburg P.H., Martial J.A., Baxter J.D., Goodman H.M.;

RT \*Construction and analysis of recombinant DNA for human chorionic  
 KT somatomammotropin";  
 RL Nature 270:494-499(1977).  
 RN [8]  
 RN SEQUENCE OF 27-217.  
 RX MEDLINE-73201971; PubMed-4712450;  
 RA Li C.H., Dixon J.S., Chung D.;  
 RT "Amino acid sequence of human chorionic somatomammotropin";  
 RL Arch. Biochem. Biophys. 155:95-110(1973).  
 RN [9]  
 RN SEQUENCE OF 27-117.  
 RX MEDLINE-72016313; PubMed-5286363;  
 RA Sherwood L.M., Handwerger S., McLaurin W.D., Lanner M.;  
 RT "Amino acid sequence of human placental lactogen";  
 RL Nature New Biol. 233:59-61(1971).  
 RN [10]  
 RN ERRATUM.  
 RA Sherwood L.M., Handwerger S., McLaurin W.D., Lanner M.;  
 RL Nature New Biol. 235:64-64(1972).  
 RN [11]  
 RN INTERCHAIN DISULFIDE BONDS.  
 RX MEDLIN -79173081; PubMed-438159;  
 RA Schnelzer A.B., Kowalski K., Russell J., Sherwood L.M.;  
 RT "Identification of the inter chain disulfide bonds of dimeric human  
 RL placental lactogen";  
 RL J. Biol. Chem. 254:3782-3787(1979).  
 CC -|- FUNCTION: SIMILAR TO THAT OF SOMATOTROPIN.  
 CC -|- SUBCELLULAR LOCATION: Secreted.  
 CC -|- MISCELLANEOUS: THE SEQUENCE OF CSH1 IS SHOWN.  
 CC -|- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.  
 CC .....  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC .....  
 DR EMBL: V00573; CAA23836.1; -  
 DR EMBL: J00289; AAA98747.1; -  
 DR EMBL: K02401; AAA52115.1; -  
 DR EMBL: M15894; AAA52116.1; -  
 DR EMBL: J03071; AAA52551.1; -  
 DR EMBL: J00118; AAA98621.1; -  
 DR EMBL: BC002717; AAH02717.1; -  
 DR EMBL: BC005921; AAH05921.1; -  
 DR EMBL: BC020756; AAH20756.1; -  
 DR PIR: A26449; A26449.  
 DR PIR: G32435; LCHUC.  
 DR HSSP: P01241; 1A22.  
 DR Genew: HGNC:2440; CSH1.  
 DR Genew: HGNC:2441; CSH2.  
 DR MIN: I50200;  
 DR GO: GO:0007565; P:pregnancy; TAS.  
 DR InterPro: IPR001400; Somatotropin.  
 DR Pfam: PF00103; hormone; 1.  
 DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
 DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
 DR Hormone; Placenta; Multigene family; Signal.  
 KW SIGNAL 1 26  
 FT CHAIN 27 217 LACTOGEN.  
 FT DISULFID 79 191  
 FT DISULFID 208 215  
 FT DISULFID 208 208  
 FT DISULFID 215 215  
 FT VARIANT 3 3  
 FT VARIANT 104 105  
 FT CONFLICT 84 84  
 FT CONFLICT 95 95  
 FT CONFLICT 116 116

```

db      27  FPAMPSSLSFANAVLRAQHLHQAADTKFERAYIPEGORYS-IONAQTATCFSETIPIA 81
      QY      62  PSNREETOQKSNLELLRISLLIQSWLEPVQ 92
      DB      86  PTGKEAQQRSDMELLRFSLLLIQSWLGPVQ 116

RESULT 11
SOMA_BALBO
ID SOMA_BALBO STANDARD; PRT: 190 AA.
AC P33092;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Somatotropin (Growth hormone).
DI GHI.
GN Balaenoptera borealis (Sei whale).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;
OC Balaenopteridae; Balaenoptera.
ON NCBI_TaxID-9768;
RX 11)
RP SEQUENCE.
RX MEDLINE-83000569; PubMed-7115813;
RX Yudaev N.A., Pankov Y.A., Bulatov A.A., Osipova T.A.;
RT "Amino acid sequence of seiwhale somatotropin.";
RT Biochimia 47:1059-1069(1982).
RN [2]
RP PRELIMINARY PARTIAL SEQUENCE.
RA Osipova T.A., Bulatov A.A., Pankov Y.A.;
RT "Structural studies of tryptic peptides from large cyanogen bromide
RT fragments of sei whale (Balaenoptera borealis) somatotropin.";
RT Biorg. Khim. 4:1589-1599(1978).
CC !- FUNCTION: Plays an important role in growth control. Its major
CC role in stimulating body growth is to stimulate the liver and
CC other tissues to secrete IGF-1. It stimulates both the
CC differentiation and proliferation of myoblasts. It also stimulates
CC amino acid uptake and protein synthesis in muscle and other
CC tissues.
CC !- SUBCELLULAR LOCATION: Secreted.
CC !- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.
DR PIR: P01440; P01440.
DR HSP: P01241; IAXI.
DR InterPro: IPR0C1400; Somatotropin.
DR Pfam: PF00103; Hormone; 1.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
KW Hormone; Pituitary.
KW FTDISULFID 52 163 HY SIMILARITY.
FT DISULFID 180 188 HY SIMILARITY.
SQ SEQUENCE 190 AA; 21835 MW; 09F6FF6DE1A475D6 CRC64;

Query Match 38.6%; Score 307.5; DB 1; Length 190;
Best Local Similarity 67.0%; Pred. No. 4.3e-23;
Matches 61; Conservative 14; Mismatches 15; Indels 1; Gaps 19

QY 2 FPTPLSLRLFDNMLRAHLHLQAFDTYQEFEEAYIPKEQYKSFLONPOTLSLFSFSIPT 61
DB 1 FPMPLSSLSFANAVLRAQHLHQAADTKFERAYIPEGORY-FLONAQSTGCFSEVPI 59

QY 62 PSNREETOQKSNLELLRISLLIQSWLEPVQ 92
DB 60 PANDEAQQRSDVELLRFSLLLIQSWLGPVQ 90

RESULT 12
SOMA_HORSE
ID SOMA_HORSE STANDARD; PRT: 216 AA.
AC P01245;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)

```

```

DE GN Somatotropin precursor (Growth hormone).
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1];
RP SEQUENCE FROM N.A.
RC TISSUE=Plutitary; PubMed=8206392;
RA MEDLINE=94266171; PubMed=8206392;
RA Ascacio-Martinez J.A., Barrera-Saldana H.A.;
RT "Sequence of a cDNA encoding horse growth hormone.";
RL Gene 143:299-300(1994).
RN [2];
RP SEQUENCE OF 27-216.
RX MEDLINE=77005410; PubMed=965151;
RA Zakin M.M., Poskus E., Langton A.A., Ferrara P., Santome J.A.,
RA Dellacha J.M., Paladini A.C.;
RT "Primary structure of equine growth hormone.";
RL Int. J. Pept. Protein Res. 8:435-444(1976).
RN [3];
RP PRELIMINARY SEQUENCE OF 27-216.
RX MEDLINE=74020362; PubMed=4747849;
RA Zakin M.M., Poskus E., Dellacha J.M., Paladini A.C., Santome J.A.;
RT "The amino acid sequence of equine growth hormone.";
RL FEBS Lett. 34:353-355(1973).
RN [4];
RP SEQUENCE OF 68-95 AND 183-216.
RA Zakin M.M., Poskus E., Dellacha J.M., Paladini A.C., Santome J.A.;
RT "Amino acid sequences around the cystine residues in equine growth hormone.";
RL FEBS Lett. 25:77-82(1972).
RN [5];
RP SEQUENCE OF 202-216.
RA Oliver L., Hartree A.S.;
RT "Amino acid sequences around the cystine residues in horse growth hormone.";
RL Biochem. J. 109:19-24(1968).
CC -!- FUNCTION: Plays an important role in growth control. Its major
CC role in stimulating body growth is to stimulate the liver and
CC other tissues to secrete IGF-1. It stimulates both the
CC differentiation and proliferation of myoblasts. It also stimulates
CC amino acid uptake and protein synthesis in muscle and other
CC tissues.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U02929; AAA21027.1; -
CC HSSP: P01246; 1BS1.
CC InterPro: IPR001400; Somatotropin.
CC Pfam: PF00103; hormone;
CC PRINTS: PR00836; SOMATOTROPIN.
CC PROSITE: PS00266; SOMATOTROPIN_1; 1.
CC PROSITE: PS00338; SOMATOTROPIN_2; 1.
CC Hormone; Pituitary; Signal.
CC SIGNAL 1 26
CC CHAIN 27 216 SOMATOTROPIN.
CC DISULFID 78 189 BY SIMILARITY.
CC DISULFID 206 214 BY SIMILARITY.
CC SEQUENCE 216 AA; 24423 MW; 37AB3173834D11AC CRC64;
CC
CC Query Match 38.5%; Score 306.5; DB 1; Length 216;
CC Best Local Similarity 65.2%; Pred. No. 6.2e-23;
CC Matches 60; Conservative 14; Mismatches 17; Indels 1; Gaps 1;
CC
DE GN Somatotropin precursor (Growth hormone).
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1];
RP SEQUENCE FROM N.A.
RC TISSUE=Plutitary; PubMed=8206392;
RA MEDLINE=94266171; PubMed=8206392;
RA Ascacio-Martinez J.A., Barrera-Saldana H.A.;
RT "Sequence of a cDNA encoding horse growth hormone.";
RL Gene 143:299-300(1994).
RN [2];
RP SEQUENCE OF 27-216.
RX MEDLINE=77005410; PubMed=965151;
RA Zakin M.M., Poskus E., Langton A.A., Ferrara P., Santome J.A.,
RA Dellacha J.M., Paladini A.C.;
RT "Primary structure of equine growth hormone.";
RL Int. J. Pept. Protein Res. 8:435-444(1976).
RN [3];
RP PRELIMINARY SEQUENCE OF 27-216.
RX MEDLINE=74020362; PubMed=4747849;
RA Zakin M.M., Poskus E., Dellacha J.M., Paladini A.C., Santome J.A.;
RT "The amino acid sequence of equine growth hormone.";
RL FEBS Lett. 34:353-355(1973).
RN [4];
RP SEQUENCE OF 68-95 AND 183-216.
RA Zakin M.M., Poskus E., Dellacha J.M., Paladini A.C., Santome J.A.;
RT "Amino acid sequences around the cystine residues in equine growth hormone.";
RL FEBS Lett. 25:77-82(1972).
RN [5];
RP SEQUENCE OF 202-216.
RA Oliver L., Hartree A.S.;
RT "Amino acid sequences around the cystine residues in horse growth hormone.";
RL Biochem. J. 109:19-24(1968).
CC -!- FUNCTION: Plays an important role in growth control. Its major
CC role in stimulating body growth is to stimulate the liver and
CC other tissues to secrete IGF-1. It stimulates both the
CC differentiation and proliferation of myoblasts. It also stimulates
CC amino acid uptake and protein synthesis in muscle and other
CC tissues.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U02929; AAA21027.1; -
CC HSSP: P01246; 1BS1.
CC InterPro: IPR001400; Somatotropin.
CC Pfam: PF00103; hormone; 1
CC PROSITE: PS00266; SOMATOTROPIN_1; 1.
CC PROSITE: PS00338; SOMATOTROPIN_2; 1.
CC Hormone; Pituitary; Signal.
CC SIGNAL 1 26
CC CHAIN 27 216 SOMATOTROPIN.
CC DISULFID 78 189
CC DISULFID 206 214
CC SEQUENCE 216 AA; 24423 MW; 37AB3173834D11AC CRC64;
CC
CC Query Match 38.5%; Score 306.5; DB 1; Length 216;
CC Best Local Similarity 65.2%; Pred. No. 6.2e-23;
CC Matches 60; Conservative 14; Mismatches 17; Indels 1; Gaps 1;

```

RESULT 14

```

RESOL1 14
SOMA_NICPY STANDARD: PRT: 217 AA.
AC Q9GMB2;
AC Q9GMB2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Somatotropin precursor (Growth hormone).
GN GHI.
OS Nycticebus pygmaeus (Pygmy slow loris).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Strepsirrhini; Loridae; Nycticebus.
OC NCBI_TaxID=101278;
RN [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Liver;
RC Wallis O.C.; Zhang Y.P.; Wallis M.;
RA "Cloning and characterization of the gene encoding slow loris growth
RT hormone."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Plays an important role in growth control. Its major
CC role in stimulating body growth is to stimulate the liver and
CC other tissues to secrete IGF-1. It stimulates both the
CC differentiation and proliferation of myoblasts. It also stimulates
CC amino acid uptake and protein synthesis in muscle and other
CC tissues.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).

```

```

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85261358; PubMed=2991252;
RA Linzer D.I.H., Talamantes F.;
RT "Nucleotide sequence of mouse prolactin and growth hormone mRNAs and
RT expression of these mRNAs during pregnancy.";
RN J. Biol. Chem. 260:9574-9574(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=E2TD; TISSUE=Liver;
RX MEDLINE=96194803; PubMed=86
RA Das P., Meyer L., Seyfert H.;
RT "Structure of the growth hormone-encoding gene and its promoter in
RT mice.";
RN Gene 169:209-213(1996).
RP -!- FUNCTION: Plays an important role in growth control. Its major
RT role in stimulating body growth is to stimulate the liver and
CC other tissues to secrete IGF-1. It stimulates both the
CC differentiation and proliferation of myoblasts. It also stimulates
CC amino acid uptake and protein synthesis in muscle and other
CC tissues.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X02891; CAA36650.1;
DR ENBL: 246663; CAA86658.1;
DR PIR: B23911; STMS.
DR HSP: P01246; 1HST.
DR MGD: MGI:95707; Gh.
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone_1
DR PROSITE: PS00266; SOMATOTROPIN_1;
DR PROSITE: PS00338; SOMATOTROPIN_2;
DR Hormone: Pituitary; Signal.
DR SIGNAL 1 26 BY SIMILARITY.
DR CHAIN 27 216 SOMATOTROPIN.
DR DISULFID 78 184 PY SIMILARITY.
DR DISULFID 206 214 HY SIMILARITY.
DR SEQUENCE 216 AA; 24716 MW; 98656A7AE25D65FC CRC64;
Query Match 38.28; Score 304.5; DB 1; Length 216;
Best Local Similarity 64.88; Pred. No. 9.7e-23;
Matches 59; Conservative 14; Mismatches 17; Indels 1; Gaps 1;
QY 2 FPTPIPSLRFDNLRAHRLHQLAFDTQCFEEAYIPKEQYFLQNPOTSLSFSESIP 61
DB 27 FPAMPPLSSFLSNVLAHQHLHQLAADTYKEFRAYIPEGORYS-IQNQAQAFCSFTIPA 85
QY 62 PSNREETQOKSNLELRLISLLLTQSLPEVQ 92
DB 86 PTKKEAQORTDMELLRFSLLLLTQSLGPVQ 116

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2003, 12:34:00 : Search time 36.1868 Seconds  
(without alignments)  
1069.670 Million cell updates/sec

Title: US-09-423-100-7

Perfect score: 797

Sequence: 1 NEPTPLSLFDNMLRAHR.....IVEQCCTICSGLYLENYCN 150

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_23:\*

- 1: sp\_archea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriaph:\*
- 17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	436	54.7	217	6 Q8WNE0	Q8WNE0 ateles geof
2	427.5	53.6	245	4 O14644	O14644 homo sapien
3	407.5	51.1	217	6 Q07369	Q07369 macaca mula
4	396	49.7	212	6 Q07368	Q07368 macaca mula
5	396	49.7	217	6 Q07367	Q07367 macaca mula
6	381	47.8	217	4 Q14407	Q14407 homo sapien
7	348	43.7	217	6 Q8WNE9	Q8WNE9 ateles geof
8	341	42.8	202	4 O14643	O14643 homo sapien
9	322.5	40.5	217	6 Q8M174	Q8M174 callithrix
10	306.5	38.5	216	11 Q70615	Q70615 spalax leuc
11	301.5	37.8	216	6 Q8M173	Q8M173 delphinus d
12	301.5	37.8	216	6 Q8HYE5	Q8HYE5 ailuropoda
13	298.5	37.5	216	11 Q9R2C3	Q9R2C3 mus musculu
14	297.5	37.3	204	6 Q95205	Q95205 ovis aries
15	297.5	37.3	216	11 Q9JMK4	Q9JMK4 cavia porce
16	297	37.3	217	6 Q8M175	Q8M175 callithrix

17	290.5	36.4	192	6 Q9TU21	Q9TU21 capra hircu
18	289.5	36.3	192	6 Q9TQW9	Q9TQW9 bos indicus
19	287.5	36.1	190	11 Q9JMG0	Q9JMG0 cavia porce
20	286.5	35.9	178	6 Q95M35	Q95M35 tarsius ban
21	285.5	35.8	217	6 Q9BEC0	Q9BEC0 traquilus ja
22	285.5	35.8	217	6 Q9BEB9	Q9BEB9 traquilus ja
23	285	35.8	167	4 Q78451	Q78451 homo sapien
24	283.5	35.6	178	6 Q95M36	Q95M36 tarsius syr
25	280.5	35.2	217	6 Q28957	Q28957 sus scrofa
26	271.5	34.1	110	6 Q8HXV2	Q8HXV2 pongo pygma
27	266.5	33.4	110	11 Q91X13	Q91X13 spermophilu
28	265.5	33.3	143	6 Q95240	Q95240 canis famill
29	261.5	32.8	218	13 Q9PU72	Q9PU72 cynops pyrr
30	254	31.9	110	6 Q8WNW6	Q8WNW6 felis silve
31	246.5	30.9	145	6 Q9BDR4	Q9BDR4 galago cras
32	234	29.4	199	4 Q14406	Q14406 homo sapien
33	233.5	29.3	195	13 Q91386	Q91386 amia calva
34	224.5	28.2	106	13 Q91807	Q91807 rana pipien
35	201.5	25.3	110	13 Q98TA8	Q98TA8 pantodon bu
36	197	24.7	108	13 Q9DDE5	Q9DDE5 brachydantio
37	195.5	24.5	108	13 Q90ZM4	Q90ZM4 catla catla
38	195	24.5	111	13 Q98TB0	Q98TB0 chitala chi
39	193.5	24.3	110	13 Q90ZY1	Q90ZY1 hiodon alos
40	191.5	24.0	111	13 Q98TA7	Q98TA7 osteoqllossu
41	187.5	23.5	87	13 Q98TA9	Q98TA9 gnathonemus
42	187.5	23.5	108	13 Q98TB1	Q98TB1 catostomus
43	186.5	23.4	93	6 Q8HXV8	Q8HXV8 bos mutus g
44	185.5	23.3	91	13 Q98TB2	Q98TB2 ambloplites
45	177.5	22.3	209	13 Q8AXX9	Q8AXX9 anquilla an

#### ALIGNMENTS

RESULT 1

Q8WNE0 PRELIMINARY, PRT: 217 AA.

AC Q8WNE0:

DT 01-MAR-2002 (TRENBLrel. 20, Created)

DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)

DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)

DE Growth hormone.

GN GH-N.

OS Ateles geoffroyi (Black-handed spider monkey).

OC Eukaryota; Metazoa; Chordata; Claditae; Vertebrata; Euteleostomi;

SC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Ateles; Ateles.

OX NCBI:TaxID:9599;

RN [1]

RP SEQUENCE FROM N.A.

RA Revol A., Esquivel D., Santiago D., Barrera-Saldana H.,

RT \*Independent duplication of the growth hormone gene in three

RT Anthropoidean lineages.\*

RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF34234; AAL72286.1; -

DR InterPro: IPR001400; Somatotropin.

DR Pfam: PF00103; hormone; 1.

DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.

DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.

SO SEQUENCE 217 AA; 24894 MW; 425829FF41EEAE6 CRC64;

Query Match 54.7%; Score 436; DB 6; Length 217;  
Best Local Similarity 66.9%; Pred. No. 1.4e-38;  
Matches 97; Conservative 8; Mismatches 24; Indels 16; Gaps 4;

QY	2	FPTPLSLFDNMLRAHRLHQLAFDYVOEFEEAYIPKEOKYSFLONPOTSLSPSESPT 61		
Db	27	FPTPLSLFDNMLRAHRLHQLAFDYVOEFEEAYIPKEOKYSFLONPOTSLSPSESPT 86		
QY	62	PSNREETOQKSNLELLRLISLLIQSWLEPVLGTGPRFVNOHLCG-----SHLVE 111		
Db	87	PASKKETQKSNLELLRLISLLIQSWLEPVLGTGPRFVNOHLCG-----SHLVE 145		
QY	112	ALYLVCG--ERGFYTPKTRGIVEQ 134		

Db 146 GIOTLMORLEDG---SPUTGEIFRQ 167

## RESULT 2

014644  
ID O14644 PRELIMINARY; PRT: 245 AA.  
AC O14644; 1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Placental growth hormone isoform hGH-V3 precursor.  
GN HGH-V.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Term placenta;  
RA MEDLINE=94373737; PubMed=9709963;  
RX Boguszewski C.L., Svensson P.A., Jansson T., Clark R.,  
RA Carlsson L.M.S., Carlsson B.;  
RT \*Cloning of two novel growth hormone transcripts expressed in human  
RT placenta\*;  
RL J. Clin. Endocrinol. Metab. 83:2878-2885(1998).  
RL EMBL: AF006061; AAB71829.1; -;  
DR HSSP: P01241; IAZ2.  
DR InterPro: IPR001400; Somatotropin.  
DR Pfam: PF00103; hormone; 1.  
DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
KW Signal.  
FT SIGNAL 1 26 POTENTIAL  
SQ SEQUENCE 245 AA; 27101 MW; 14CC7F8CD75D91C8 CRC64;

Query Match 53.6%; Score 427.5; DB 4; Length 245;  
Best Local Similarity 71.9%; Pred. No. 1.3e-37;  
Matches 92; Conservative 6; Mismatches 17; Indels 13; Gaps 2;  
QY 2 FPTPLSLFDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONPOTSLSFSESIPTP 61  
DB 27 FPTPLSLFDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONPOTSLSFSESIPTP 86  
QY 62 PSNREETQOKSNLELLRISLLIQSWLEPVQ-----GTGPRFVNOHLCGSHIV 110  
DB 87 PSNREETQOKSNLELLRISLLIQSWLEPVQ...KSVFANSVLVYGASDSNVYRHL---KDLF 144  
QY 111 EALYLVCG 118  
DB 145 EGIOTLLIC 152

## RESULT 3

Q07369  
ID Q07369 PRELIMINARY; PRT: 217 AA.  
AC Q07369;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Chorionic somatomammotropin-3.  
OS Macaca mulatta (Rhesus macaque).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;  
OC Cercopitheidae; Macaca.  
OX NCBI\_TaxID=9544;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Midpregnancy placenta;  
RX MEDLINE=94008724; PubMed=8404617;  
RA Golos T.G., Durning M., Fisher J.M., Fowler P.D.;  
RT \*Cloning of four growth hormone/chorionic somatomammotropin-related  
RT complementary deoxyribonucleic acids differentially expressed during  
RT pregnancy in the rhesus monkey placenta\*;  
RT

RL Endocrinology 133:1744-1752(1993).  
DR EMBL: L16554; AAA18641.1; -;  
DR HSSP: P01241; IAXI.  
DR InterPro: IPR001400; Somatotropin.  
DR Pfam: PF00103; hormone; 1.  
DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
DR PROSITE: PS00138; SOMATOTROPIN\_2; 1.  
SQ SEQUENCE 217 AA; 24874 MW; F1E66AFDBBA1B185 CRC64;

Query Match 51.1%; Score 407.5; DB 6; Length 217;  
Best Local Similarity 71.9%; Pred. No. 1.6e-35;  
Matches 82; Conservative 12; Mismatches 19; Indels 1; Gaps 1;

QY 3 PTPLSLFDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONPOTSLSFSESIPTP 62  
DB 28 PSVPLSLFDNIMQAHRLHQLAFDTYQFEFEAYIPKEKKHSLMGNPQASFCFESIPTP 87  
QY 63 SNREETQOKSNLELLRISLLIQSWLEPVQGTGPRFVNOHLCGSHIV 116  
DB 88 SNREETQOKSNLELLRISLLIQSWLEPVQ-LGSVFANLVYGTSESAYDGL 140

## RESULT 4

Q07368  
ID Q07368 PRELIMINARY; PRT: 212 AA.  
AC Q07368;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Somatotropin 2 precursor (Growth hormone 2) (Fragment).  
OS Macaca mulatta (Rhesus macaque).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;  
OC Cercopitheidae; Macaca.  
OX NCBI\_TaxID=9544;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=94008724; PubMed=8404617;  
RA Golos T.G., Durning M., Fisher J.M., Fowler P.D.;  
RT \*Cloning of four growth hormone/chorionic somatomammotropin-related  
RT complementary deoxyribonucleic acids differentially expressed during  
RT pregnancy in the rhesus monkey placenta\*;  
RL Endocrinology 133:1744-1752(1993).  
DR EMBL: L16553; AAA18840.1; -;  
DR HSSP: P01241; IAXI.  
DR InterPro: IPR001400; Somatotropin.  
DR Pfam: PF00103; hormone; 1.  
DR PROSITE: PS00138; SOMATOTROPIN\_2; 1.  
FT NON\_TER 1  
SQ SEQUENCE 212 AA; 24526 MW; 27B0941106256EAF5 CRC64;

Query Match 49.7%; Score 396; DB 6; Length 212;  
Best Local Similarity 82.2%; Pred. No. 2.6e-34;  
Matches 74; Conservative 11; Mismatches 5; Indels 0; Gaps 0;

QY 3 PTPLSLFDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONPOTSLSFSESIPTP 62  
DB 23 PSVPLSLFDHAMIQAHLHQLAFDTYQFEFEAYIPKEKKHSLMGNPQASFCFESIPTP 82  
QY 13 SNREETQOKSNLELLRISLLIQSWLEPVQ 92  
DB 83 SNLEETQOKSNLELLRISI-LIQSWLEPVQ 112

## RESULT 5

Q07367  
ID Q07367 PRELIMINARY; PRT: 217 AA.  
AC Q07367;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Chorionic somatomammotropin-1.

```

OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Midpregnancy placenta.
RX MEDLINE=94008724; PubMed=8404617;
RA Golos T.G., Durning M., Fisher J.M., Fowler P.D.:
RT "Cloning of four growth hormone/chorionic somatomammotropin-related
RT complementary deoxyribonucleic acids differentially expressed during
RT pregnancy in the rhesus monkey placenta.";
RL Endocrinology 133:1744-1752(1993).
DR EMBL: L16552; AAA18839.1; -
DR HSSP: P01241; JAXI
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone; 1.
DR PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
SQ SEQUENCE 217 AA: 24942 MW: 55288.915131F2BC CRC64:

Query Match 49.7%; Score 396; DB 6; Length 217;
Best Local Similarity 82.2%; Pred. No. 2.7e-34;
Matches 74; Conservative 11; Mismatches 5; Indels 0; Gaps 0;

QY 3 PTIPSLRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPQTSLSFSSEIPTP 62
DL PVPVLSRLFDHAMIQAHRHQLAFDTYQEFEEAYIPKEKHSLSMENPQASFCFADSIPTP 87
QY 63 SNRETOOKSNLELLRLSLLIQSMLPQVQ 92
DB SLEETOOKSNLELLRLSLLIQSMLPQVQ 117

RESULT 6
Q14407
ID Q14407 PRELIMINARY; PRT: 217 AA.
AC Q14407;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Chorionic somatomammotropin CS-2 (Chorionic somatomammotropin hormone
DE 2).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89307277; PubMed=2744760;
RA Chen E.Y., Liao Y.C., Smith D.H., Barrera-Saldana H.A., Gelin R.E.,
RA Seeburg P.H.:
RT "The human growth hormone locus: nucleotide sequence, biology, and
RT evolution.";
RL Genomics 4:479-497(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91102558; PubMed=1980158;
RA Vnencak-Jones C.L., Phillips J.A., III.:
RT "Hot spots for growth hormone gene deletions in homologous regions
RT outside of Alu repeats.";
RL Science 250:1745-1748(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.:
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.:
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.

```

```

DR EMBL: J03071; AAA52553.1; -
DR EMBL: BC022044; AAH22044.1; -
DR EMBL: BC035965; AAH35965.1; -
DR HSSP: P01241; IA22.
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone; 1.
DR PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
SQ SEQUENCE 217 AA: 24994 MW: 39FAACDDH6B2E95; CRC64:

Query Match 47.8%; Score 381; DB 4; Length 217;
Best Local Similarity 82.0%; Pred. No. 1.1e-32;
Matches 73; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

QY 4 TIPLSKLFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPQTSLSFSSEIPTP 63
DB TVPLSKLFDHAMIQAHRHQLAFDTYQEFEEAYIPKQKYSFLHRSQTSFQFSISPTPS 88
QY 64 NREETOOKSNLELLRLSLLIQSMLPQVQ 92
DB NMEETOOKSNLELLRLSLLIQSMLPVR 117

RESULT 7
Q8WND9
ID Q8WND9 PRELIMINARY; PRT: 217 AA.
AC Q8WND9;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Growth hormone.
GN GH-V.
OS Ateles geoffroyi (Black-handed spider monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Ateles.
OX NCBI_TaxID=9509;
RN [1]
RP SEQUENCE FROM N.A.
RA Revol A., Esquivel B., Santiago B., Barrera-Saldana H.:
RT "Independent duplication of the growth hormone gene in three
RT Anthropoid lineages.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF374235; AAL72287.1; -
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone; 1.
DR PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
SQ SEQUENCE 217 AA: 25293 MW: 74174.5ALH57053E CRC64:

Query Match 43.7%; Score 348; DB 6; Length 217;
Best Local Similarity 75.8%; Pred. No. 3.7e-29;
Matches 69; Conservative 8; Mismatches 14; Indels 0; Gaps 0;

QY 2 FTIPSLRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPQTSLSFSSEIPT 61
DB FPRIPSLRFDGAMLRHQLHQLAFDTYQELLENCIPKQKQYFELRNKPNLFCSESIPT 86
QY 62 PSNREETOOKSNLELLRLSLLIQSMLPQVQ 92
DB PFNKEVLAKSLELLHIS LLIQSMLPQV 117

RESULT 8
O14643
ID O14643 PRELIMINARY; PRT: 202 AA.
AC O14643;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Placental growth hormone 20kDa isoform precursor.
GN HGH-V

```

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Term placenta;  
 RX MEDLINE=98373737; PubMed=9709963;  
 RA Boguszewski C.L., Svensson P.A., Jansson T., Clark R.,  
 Carlsson L.M.S., Carlsson B.,  
 RT "Cloning of two novel growth hormone transcripts expressed in human  
 placenta.";  
 RL J. Clin. Endocrinol. Metab. 83:2878-2885(1998).  
 DR EMBL: AF006060; AAB71828.1; -;  
 DR HSSP: P01241; 1A22  
 DR InterPro: IPR001400; Somatotropin.  
 DR Pfam: PF00103; hormone\_1; Somatotropin.  
 DR PRINTS: PR00836; SOMATOTROPIN.  
 DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
 DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
 DR SIGNAL.  
 KW SIGNAL.  
 FT SIGNAL  
 SQ SEQUENCE 202 AA; 23128 MW; 38B64D01A9197C6 CRC64;  
 Query Match 42.8%; Score 341; DB 4; Length 202;  
 Best Local Similarity 65.8%; Pred. No. 1.9e-28;  
 Matches 75; Conservative 4; Mismatches 9; Indels 26; Gaps 2;  
 QY 2 FPTPLSLRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLONPQTSLSFSISPT 61  
 DB 27 FPTPLSLRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLONPQTSLSFSISPT 71  
 QY 62 PSNRETOQKSNLELLRISLLIQSWLEPVQL-----GTGPRFVNHQL 104  
 DB 72 PSNRETOQKSNLELLRISLLIQSWLEPVQLRVSFANSVYRHL 125  
 RESULT 9  
 QY 62 PSNRETOQKSNLELLRISLLIQSWLEPVQL-----GTGPRFVNHQL 104  
 DB 72 PSNRETOQKSNLELLRISLLIQSWLEPVQLRVSFANSVYRHL 125  
 ID Q8MI74 PRELIMINARY; PRT: 217 AA.  
 AC Q8MI74;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Growth hormone-like protein 6 precursor.  
 GN GHLF6.  
 OS Callithrix jacchus (Common marmoset).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Callithrix.  
 OX NCBI\_TaxID=9483;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC "Characterisation of the GH gene cluster in a new-world monkey, the  
 marmoset (Callithrix jacchus).";  
 RL J. Mol. Endocrinol. 0:0-0(2002).  
 DR EMBL: AJ489811; CAD34012.1; -;  
 DR InterPro: IPR001400; Somatotropin.  
 DR Pfam: PF00103; hormone\_1;  
 DR PRINTS: PR00836; SOMATOTROPIN.  
 DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
 DR SIGNAL.  
 KW SIGNAL.  
 FT CHAIN  
 SQ SEQUENCE 217 AA; 25177 MW; 5ECF148798278F1A CRC64;  
 Query Match 40.5%; Score 322.5; DB 6; Length 217;  
 Best Local Similarity 64.7%; Pred. No. 2e-26;  
 Matches 66; Conservative 13; Mismatches 22; Indels 1; Gaps 1;  
 QY 3 PTPPLSLRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLONPQTSLSFSISPT 62  
 DB 28 PTPPLSLRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLONPQTSLSFSISPT 87

QY 63 SNRETOQKSNLELLRISLLIQSWLEPVQLGTGPRFVNHQL 104  
 DB 88 FHKREMLGKSNVLELLHISLLIQSWLEPVQLGTGPRFVNHQL 128  
 RESULT 10  
 QY 63 SNRETOQKSNLELLRISLLIQSWLEPVQLGTGPRFVNHQL 104  
 DB 88 FHKREMLGKSNVLELLHISLLIQSWLEPVQLGTGPRFVNHQL 128  
 ID Q70615 PRELIMINARY; PRT: 216 AA.  
 AC Q70615;  
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Growth hormone precursor  
 OS Spalax leucodon ehrenbergi (Ehrenberg's mole rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Spalacinae;  
 OC Nannospalax.  
 OX NCBI\_TaxID=30637;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99124645; PubMed=9742177;  
 RA Lioupi A., Nevo E., Wallis M.;  
 RT "Cloning and characterisation of the gene encoding mole rat (Spalax  
 ehrenbergi) growth hormone.";  
 RL J. Mol. Endocrinol. 22:29-35(1999).  
 DR EMBL: AJ005819; CAA06716.1; -;  
 DR HSSP: P01241; 1A21.  
 DR InterPro: IPR001400; Somatotropin.  
 DR Pfam: PF00103; hormone\_1;  
 DR PRINTS: PR00836; SOMATOTROPIN.  
 DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
 DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
 DR SIGNAL.  
 KW SIGNAL.  
 FT CHAIN  
 SQ SEQUENCE 216 AA; 24627 MW; FEAB8A523BA0ADFE CRC64;  
 Query Match 38.5%; Score 306.5; DB 11; Length 216;  
 Best Local Similarity 65.9%; Pred. No. 1e-24;  
 Matches 60; Conservative 13; Mismatches 17; Indels 1; Gaps 1;  
 QY 2 FPTPLSLRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLONPQTSLSFSISPT 61  
 DB 27 FPAAPLSNLAFANALRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLONPQTSLSFSISPT 85  
 QY 62 PSNRETOQKSNLELLRISLLIQSWLEPVQL-----GTGPRFVNHQL 92  
 DB 86 PTKREEMAGKSNVLELLHISLLIQSWLEPVQL-----GTGPRFVNHQL 116  
 RESULT 11  
 QY 62 PSNRETOQKSNLELLRISLLIQSWLEPVQL-----GTGPRFVNHQL 92  
 DB 86 PTKREEMAGKSNVLELLHISLLIQSWLEPVQL-----GTGPRFVNHQL 116  
 ID Q8MI73 PRELIMINARY; PRT: 216 AA.  
 AC Q8MI73;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Growth hormone precursor.  
 GN GH.  
 OS Delphinus delphis (saddleback dolphin) (Black sea dolphin).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti; Delphinidae;  
 OC Delphinus.  
 OX NCBI\_TaxID=9728;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RA Maniou Z., Wallis O.C., Wallis M.;  
 RT "Cloning and characterisation of the GH gene from the common dolphin  
 (Delphinus delphis).";  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ492191; CAD37292.1; -;

```

DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone; 1.
DR PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
KW Signal.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 216 GROWTH HORMONE.
SQ SEQUENCE 216 AA; 24509 MW; 1EC467A84CCFEB02 CRC64;

Query Match 37.8%; Score 301.5; DB 6; Length 216;
Best Local Similarity 64.8%; Pred. No. 3.4e-24;
Matches 59; Conservative 14; Mismatches 17; Indels 1; Gaps 1;

QY 2 FPTPLSRFLDNAMLRHQLAFDTYQFEFAYIPKEQKYSFLONPQTSLSFSSEIPT 61
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 27 FPAMPLSSLFANAVLRHQHLQAADTYKEFERAYIPKEQKYS-IONQAACFSEIIPA 85

QY 62 PSNREETOQKSNLELLRISLLIQSWLEPVQ 92
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 86 PTGKDEAQQRSDELLRISLLIQSWLGPVQ 116

RESULT 12
Q8HYE5 PRELIMINARY; PRT: 216 AA.
ID Q8HYE5
AC Q8HYE5;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DE 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Growth hormone precursor.
OS Alluropoda melanoleuca (Giant panda).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissiped; Ursidae; Alluropoda.
OX NCBI_TaxID=9646;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Pituitary.
RA Liao M., Zhu M., Zhang A.;
RT "Cloning and expression of cDNA encoding growth hormone from
   Alluropoda melanoleuca."
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF540936; AAN7228.1; -.
KW Signal.
FT SIGNAL 1 26 POTENTIAL.
SQ SEQUENCE 216 AA; 24383 MW; 44EC17EC4BCB056 CRC64;

Query Match 37.8%; Score 301.5; DB 6; Length 216;
Best Local Similarity 64.8%; Pred. No. 3.4e-24;
Matches 59; Conservative 14; Mismatches 17; Indels 1; Gaps 1;

QY 2 FPTPLSRFLDNAMLRHQLAFDTYQFEFAYIPKEQKYSFLONPQTSLSFSSEIPT 61
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 27 FPAMPLSSLFANAVLRHQLAADTYKEFERAYIPKEQKYS-IONQAACFSEIIPA 85

QY 62 PSNREETOQKSNLELLRISLLIQSWLEPVQ 92
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 86 PTGKDEAQQRSDELLRISLLIQSWLGPVQ 116

RESULT 13
Q9R2C3 PRELIMINARY; PRT: 216 AA.
ID Q9R2C3
AC Q9R2C3;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Growth hormone.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]

```

---

```

RP SEQUENCE FROM N.A.
RA Nguyen T.N.K., Liehaber S.;
RT "Mouse Growth Hormone locus Nucleotide Sequence and Phylogenetic
   Analyses."
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: U34362; AAC99988.1; -.
DR HSSP: P01241; IAXI.
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone; 1.
DR PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
SQ SEQUENCE 216 AA; 24682 MW; EC2A06DA02536B18 CRC64;

Query Match 37.5%; Score 298.5; DB 11; Length 216;
Best Local Similarity 63.7%; Pred. No. 7.2e-24;
Matches 58; Conservative 14; Mismatches 18; Indels 1; Gaps 1;

QY 2 FPTPLSRFLDNAMLRHQLAFDTYQFEFAYIPKEQKYSFLONPQTSLSFSSEIPT 61
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 27 FPAMPLSSLFANAVLRHQLAADTYKEFERAYIPKEQKYS-IONQAACFSEIIPA 85

QY 62 PSNREETOQKSNLELLRISLLIQSWLEPVQ 92
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 86 PTGKEAQQRTMELLRISLLIQSWLGPVQ 116

RESULT 14
Q95205 PRELIMINARY; PRT: 204 AA.
ID Q95205
AC Q95205;
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Placental growth hormone (Fragment).
DE Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-placenta;
RX MEDLINE=97050633; PubMed=8673361;
RA Lacroix M.C., Devinoy E., Javelly J.L., Puissant C., Kann G.;
RT "Expression of the growth hormone gene in ovine placenta: detection
   and cellular localization of the protein."
RL Endocrinology 137:4886-4891, 1996.
DR EMBL: U49063; AAC48679.1; -.
DR HSSP: P01241; IAXI.
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone; 1.
DR PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
FT NON_TER 1
SQ SEQUENCE 204 AA; 23462 MW; EBB9451892635C6 CRC64;

Query Match 37.3%; Score 297.5; DB 6; Length 204;
Best Local Similarity 64.8%; Pred. No. 8.6e-24;
Matches 59; Conservative 14; Mismatches 17; Indels 1; Gaps 1;

QY 2 FPTPLSRFLDNAMLRHQLAFDTYQFEFAYIPKEQKYSFLONPQTSLSFSSEIPT 61
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 15 FPANSLSLFANAVLRHQLAADTYKEFERAYIPKEQKYS-IONQAACFSEIIPA 73

QY 62 PSNREETOQKSNLELLRISLLIQSWLEPVQ 92
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 74 PTSKNEAQQKSDLELLRISLLIQSWLGPVQ 104

RESULT 15
Q9JKM4

```

```

ID Q9JKM4 PRELIMINARY; PRT: 216 AA.
AC Q9JKM4;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Growth hormone precursor.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RA Odorico D.M., Fuller P.J., Herington A.C.;
RT "Cloning and sequence of guinea pig growth hormone (GH).";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF233853; AAF36409.1;
DR HSSP: P01241; LAXI.
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone. 1.
DR PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
KW Signal.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 216 GROWTH HORMONE
SQ SEQUENCE 216 AA; 24822 MW; 45996BELL9B08DD3 CRC64;

Query Match 37.3%; Score 297.5; DB 11; Length 216;
Best Local Similarity 63.7%; Pred. No. 9.2e-24;
Matches 58; Conservative 14; Mismatches 18; Indels 1; Gaps 1;

QY 2 FPTPLSLFLDNLRAHLHQLAFDTYCEEEAYIPKEQKYSFLONPQTSLSFSESIPT 61
Db 27 FRAMPLSSLFGNVLRQAHLQAADTYKEFRTYIPECQKYS-IHNTQTAFCESETIPA 85

QY 62 PSNRETOOKSNLELLRLISLLLIQSWLEPVQ 92
Db 86 PTDKEAQQRSDVELLHFSLLLIQSWLGPVQ 116

```

Search completed: September 16, 2003, 12:40:01  
 Job time : 37.1868 secs